高等学校专业英语教材

电子信息工程专业英语教程（第3版）

任治刚 主编

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

高等学校专业英语教材

电子信息工程专业英语教程

(第3版)

任治刚　主编

電子工業出版社
Publishing House of Electronics Industry
北京・BEIJING

内 容 简 介

本书编者在前一版的基础上，吸取多所大学在使用本书过程中提出的诸多宝贵意见，对全书进行修订和补充。本书的主要目的是使读者掌握电子信息工程专业英语术语及用法，培养和提高读者阅读和翻译专业英语文献资料的能力。本书由10个主题单元组成，涵盖电子信息领域的主要技术分支，主要内容包括电子器件、电子电路、电子系统组件、电子系统、现代数字设计、数字信号处理、语音和音频、图像和视频、嵌入式应用、电子仪器与测量等内容。每个主题单元由3篇课文、3篇阅读材料、课文词汇、课文注释和练习组成，在书后还附有课文参考译文、练习参考答案及缩略语表。为了方便教学，本书另配有电子教案、阅读材料参考译文和授课建议，向采纳本书作为教材的教师免费提供。

本书可作为电子信息工程专业的专业英语教材，也可供从事相关专业的工程技术人员参考使用。

图书在版编目(CIP)数据

电子信息工程专业英语教程/任治刚主编. —3版. —北京：电子工业出版社，2011.7
高等学校专业英语教材
ISBN 978-7-121-14018-1

Ⅰ.①电…　Ⅱ.①任…　Ⅲ.①电子信息—英语—高等学校—教材　Ⅳ.①H31

中国版本图书馆CIP数据核字(2011)第132434号

策划编辑：杨丽娟
责任编辑：杨丽娟　特约编辑：汪荣萍
印　　刷：北京市顺义兴华印刷厂
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路173信箱　邮编　100036
开　　本：787×980　1/16　印张：28　字数：613千字
印　　次：2011年7月第1次印刷
印　　数：5 000册　定价：45.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888。

质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：(010)88258888。

前　言

电子信息工程是当今国内外发展最为迅速、技术更新最为活跃的工程领域之一。为了应对国际化竞争，学生在学习阶段就应打下坚实的基础。而专业英语的阅读和写作能力就是电子信息工程专业毕业生所应具备的一项重要能力。

本书以提高学生英语阅读和写作能力，扩展、深化学生对本学科关键技术的认识，培养具备国际竞争力的技术人才为目的；本着先进、实用的选材原则和简明、系统的组织原则，充分吸收当前最新技术成果和外语教学成果，为电子信息工程专业学生提供一个提高英语水平和专业素养的平台。

为了保证本书的先进性和实用性，所有文章均出自国外最近几年电子信息各个领域的最新教材、专著及国际著名公司网站提供的技术应用文章(详见参考文献)。在具体内容的遴选上，尽量保证学生利用既有专业知识理解课文内容，并使学生通过学习加深和扩展相关专业知识。

本书编者在前一版的基础上，吸取多所大学在使用本书过程中提出的诸多宝贵意见，对全书进行修订和补充。全书由10个主题单元组成，涵盖电子信息领域的主要技术分支。这10个主题单元分别是：电子器件、电子电路、电子系统组件、电子系统、现代数字设计、数字信号处理、语音和音频、图像和视频、嵌入式应用、电子仪器与测量。每个主题单元由3篇课文，3篇阅读材料，课文词汇，课文注释和练习组成。其中，课文侧重展示本主题领域核心或关键的技术，选取了那些能够扩展、深化学生对本学科认识的内容；而阅读材料着力介绍该主题中的实用技术、重大科技成果、前沿领域及未来前景等；课文注释旨在解决课文中英语语言难点和专业知识难点。练习则利用与单元主题相关的材料，以完型填空、英汉互译的形式培养学生使用英语的能力。在本书书后还附有课文参考译文、练习参考答案、缩略语表和参考文献。

本书由任治刚主编，参与本书编写的还有齐敬、王雅巍、胡伟东、齐铭、白冰等人。为了方便教学，本书另配有电子教案、阅读材料参考译文及授课建议，向采纳本书作为教材的教师免费提供(获取方式：登录电子工业出版社华信教育资源网 www.hxedu.com.cn 或电话联系 010-88254485 获得)。

由于水平所限，书中难免有纰漏和欠妥之处，请各位读者不吝赐教。可通过电子邮件将您发现的问题或者对本书的建议发送到 rzgteaching@hotmail.com。

编　者

Contents

Unit 1

Electronic Devices

Lesson 1 VLSI Technology

One of the key inventions in the history of electronics, and in fact one of the most important inventions over period, was the transistor. It was invented by Bell Laboratories [1] in 1948. In short, a transistor is a device that conducts a variable amount of electricity through it, depending on how much electricity is input to it. In other words, it is a digital switch. However, unlike the vacuum tube [2], it is solid state. This means that it doesn't change its physical form as it switches. There are no moving parts in a transistor.

The advantages of the transistor over the vacuum tube were enormous. Compared to the old technology, transistors were much smaller, faster, and cheaper to manufacture. They were also far more reliable and used much less power. The transistor is what started the evolution of the modern computer industry in motion [3].

The transistor was originally a single, discrete device, which you could place individually into a circuit much like any other. Today, some special-purpose transistors are still used that way. What allowed the creation of modern processors was the invention of the integrated circuit, which is a group of transistors manufactured from a single piece of material and connected together internally, without extra wiring [4]. Integrated circuits are also called ICs or chips.

A special material is used to make these integrated circuits. While most materials either insulate from electrical flow (air, glass, wood) or conduct electricity readily (metals, water), there are some that only conduct electricity a small amount, or only under certain conditions. These are called semiconductors. The most commonly used semiconductor is of course silicon.

By careful chemical composition and arrangement, it is possible to create a very small transistor directly on a layer of silicon, using various technologies to manipulate the material into the correct form. These transistors are small, fast and reliable, and use relatively little power. The first integrated circuit was invented in 1959 by Texas Instruments[5]. It contained just six transistors on a single semiconductor surface.

After the invention of the integrated circuit, it took very little time to realize the tremendous benefits of miniaturizing and integrating larger numbers of transistors into the same integrated circuit. More transistors (switches) were required in order to

implement more complicated functions. Miniaturization[6] was the key to integrating together large numbers of transistors while increasing hardware speed and keeping power consumption and space requirements manageable.

Large-scale integration ("LSI") came to refer to the creation of integrated circuits that had previously been made from multiple discrete components. These devices typically contained hundreds of transistors. Early computers were made from many of these smaller ICs connected together on circuit boards.

As time progressed after the invention of LSI integrated circuits, the technology improved and chips became smaller, faster and cheaper. Building on the success of earlier integration efforts, engineers learned to pack more and more logic into a single circuit. This effort became known as very large scale integration (VLSI). VLSI circuits can contain millions of transistors.

Originally, the functions performed by a processor were implemented using several different logic chips. Intel[7] was the first company to incorporate all of these logic components into a single chip. This was the first microprocessor, the 4004, introduced by Intel in 1971. All of today's processors are (highly advanced!) descendants of this original 4-bit CPU.

New Words

device [di'vais] *n.* 器件，装置
conduct [kən'dʌkt] *v.* 传导
discrete [dis'kriːt] *adj.* 离散的
integrated ['intigreitid] *adj.* 集成的
insulate ['insjuleit] *vt.* 绝缘，隔离
manipulate [mə'nipjuleit] *vt.* 操作，处理
implement ['implimənt] *vt.* 实现 *n.* 器具
consumption [kən'sʌmpʃn] *n.* 消耗
manageable ['mænidʒəbl] *adj.* 易处理的
component [kəm'pəunənt] *n.* 组件
incorporate [in'kɔːpəreit] *vt.* 一体化
descendant [di'sendənt] *n.* 后裔，后代

Phrases & Expressions

over (a/the) period (of…) 在某段时间内
in motion 在运转,处于活跃状态

Technical Terms

silicon ['silikən] *n.* 硅
transistor [træn'zistə] *n.* 晶体管
semiconductor ['semikən'dʌktə] *n.* 半导体
miniaturization ['miniətʃərai'zeiʃn] *n.* 缩微化
solid state 固态
IC *abbr.* Integrated Circuit 集成电路
LSI *abbr.* Large-Scale Integration 大规模集成
VLSI *abbr.* Very Large-Scale Integration 超大规模集成

Notes

1. Bell Laboratories 是指"贝尔实验室"。
2. "真空管"是一种内部气体全部或部分抽空的电子管。
3. 该句采用 what 引导的从句做表语,是对主语 the transistor 的强调。其意思相当于:The transistor started the evolution of the modern computer industry in motion. 在本句中,start 为及物动词。
4. 该句采用 what 引导的从句做主语,是对表语"the invention of the integrated circuit"的强调。
5. Texas Instruments 是指"德州仪器公司"。
6. "缩微化"是指减小元件和电路的尺寸,从而提高封装密度、降低功耗、减少信号传播延迟。
7. Intel 是指"英特尔公司"。

Lesson 2 Memory Devices

Memories can be made in mechanical, magnetic, optical, biological and electronic technologies. Examples of magnetic memories are tapes, floppy disks, hard drives and ferroelectric[1] RAMs. Examples of optical memories are CD-ROMs, rewrittable CDs. Electronic memory is used extensively in computer equipment since it is the fastest available. For applications where speed is less important, magnetic and optical technologies are often used.

All electronic memory today can be in separate IC format, module format, or can be part of an IC as a macrofunction or "cell". In the table below is an overview of some electronic memory. (See Table 2.1)

Table 2.1 Overview of some electronic memory

Type	Properties	Read/write	Non-volatile	Speed	Cost/bit
Flip-flop	One-bit register. Usually used as a basic building block in digital circuits	Yes	No	Ultra fast	Very high
Register	Set of flip-flops holding a byte, word or long word. Used in complex chips such as CPUs	Yes	No	Ultra fast	Very high
SRAM	Array of flip-flops that is addressable. Used for temporary storage of data or cache	Yes	No	Very fast	High
DRAM	Array of storage cells which is addressable. Used for main computing data storage	Yes	No	Fast	Moderate
ROM	Array of hard-wired cells that is addressable. Programming done at time of chip manufacture	No	Yes	Very Fast	Low
EEPROM	Electrically erasable programmable ROM. Number of write cycles is limited	Yes	Yes	Low	High

The flip-flop

A flip-flop is basically a bi-state circuit in which either a 0 or 1 state can resides. Because of its simplicity, the flip-flop is extremely fast. As a basic element, the flip-flop is used in digital circuits and ICs. A flip-flop will lose its state when the supply voltage

is removed. Therefore, it is volatile.

The register

A register is a set of flip-flops in parallel. Typically a register is 8,16,32 or 64 bits wide. Often a register is used to hold data, address pointers, etc. A register is volatile and very fast just like the flip-flop.

SRAM (Static Random Access Memory)

An SRAM is an array of addressable flip-flops. The array can be configured as such that the data comes out in single bit, 4 bit, 8bit, and etc. format. SRAM is simple, fast and volatile just like the flip-flop, its basic memory cell. SRAM can be found on microcontroller boards(either on or off the CPU chip), where the amount of memory required is small and it will not pay off to build the extra interface circuitry for DRAMs. In addition, SRAM is often used as cache[2] because of its high speed.

SRAM comes in many speed classes, ranging from several ns for cache applications to 200ns for low power applications. SRAM exists in both bipolar and MOS technology. CMOS[3] technology boasts the highest density and the lowest power consumption. Fast cache memory can be constructed in BiCMOS technology, a hybrid technology that uses bipolar transistors for extra drive. The fastest SRAM memories are available in ECL (Emitter Coupled Logic) bipolar technology. Because of the high power consumption, the memory size is limited in this technology.

A special case of SRAM memory is Content Addressable Memory (CAM)[4]. In this technology, the memory consists of an array of flip-flops, in which each row is connected to a data comparator. The memory is addressed by presenting data to it (not an address!). All comparators will then check simultaneously if their corresponding RAM register holds the same data. The CAM will respond with the address of the row (register) corresponding to the original data. The main application for this technology is fast lookup tables. These are often used in network routers.

DRAM (Dynamic Random Access Memory)

The word "dynamic" indicates that the data is not held in a flip-flop but rather in a storage cell. The data in a storage cell must be refreshed (read out and re-written) regularly because of leakage. The refresh time interval is usually 4 to 64 ms. The storage cell only requires one capacitor and one transistor, whereas[5] a flip-flop connected in an array requires 6 transistors. In trench capacitor memory technology,

which is used in all modern DRAMs, the transistor is constructed above the capacitor so that the space on chip is ultimately minimized. For this reason, DRAM technology has a lower cost per bit than SRAM technology. The disadvantage of the extra circuitry required for refreshing is easily offset by the lower price per bit when using large memory sizes.

DRAM memory is, just like SRAM memory constructed as an array of memory cells. A major difference between SRAM and DRAM, however, lies in the addressing technique. With an SRAM, an address needs to be presented and the chip will respond with presenting the data of the memory cell at the output, or accepting the data at the input and write it into the addressed cell. With DRAM technology, this simple approach is impossible since addressing a row of data without rewriting it will destruct all data in the row because of the dynamic nature.

ROM (Read Only Memory)

ROMs are also called mask[6]-ROMs or mask programmed ROMs. This is because a ROM needs to be programmed by setting its cells to either 0 or 1 at the time of manufacture. Usually the 0 or 1 is formed by the presence or absence of an aluminium line. This aluminium pattern is defined by a lithographic mask used in one of the last steps of manufacture. Therefore these devices are often called mask-ROMs.

The advantage of ROM is that it can be manufactured at the lowest price in high volumes. Another advantage in some applications is that it is impossible to alter the data once the chips are made, and that no further programming and testing are required. On the other hand, if the data or code must be changed this can be a small disaster. The rest of the chips will end in the dustbin and new chips will have to be made.

EEPROM (Electrically Erasable Programmable ROM)

This means that the chip can be programmed like an EPROM, but it can be erased electrically. As a result, no UV source is required. EEPROMs can be erased on a byte-by-byte basis.

New Words

mechanical [miˈkænikl] *adj*. 机械的

magnetic [mægˈnetik] *adj*. 磁的

optical [ˈɔptikəl] *adj*. 光学的

format [ˈfɔːmæt] *n.* 格式
volatile [ˈvɔlətail] *adj.* 易失的
static [ˈstætik] *adj.* 静态的
configure [kənˈfigə] *vt.* 配置,设定
boast [bəust] *v.* 夸耀
hybrid [ˈhaibrid] *adj.* 混合的
simultaneously [ˌsiməlˈteiniəsli] *adv.* 同时
corresponding [ˌkɔrisˈpɔndiŋ] *adj.* 相应的
dynamic [daiˈnæmik] *adj.* 动态的
whereas [weərˈæz] *conj.* 然而
offset [ˈɔːfset] *v.* 弥补,抵销
approach [əˈprəutʃ] *n.* 方法
aluminium [ˌæljuːˈminjəm] *n.* 铝
lithographic [ˌliθəˈgræfik] *adj.* 平版印刷的
alter [ˈɔːltə] *v.* 改变

Phrases & Expressions

in parallel 并行的,平行的
pay off 带来利益;偿清债务

Technical Terms

memory [ˈmeməri] *n.* 存储器,内存
ferroelectric [ˌferəuiˈlektrik] *adj.* 铁电的
register [ˈredʒistə] *n.* 寄存器
array [əˈrei] *n.* 阵列,数组
cache [kæʃ] *n.* 高速缓存
bipolar [baiˈpəulə] *adj.* 双极性的
drive [draiv] *n.* 驱动器
erasable [iˈreizəbl] *adj.* 可擦写的
comparator [ˈkɔmpəreitə] *n.* 比较器
leakage [ˈliːkidʒ] *n.* 泄漏
refresh [riˈfreʃ] *v.* 刷新
capacitor [kəˈpæsitə] *n.* 电容器

mask [mɑːsk] *n*. 掩模，掩码，掩蔽
pattern [ˈpætən] *n*. 模式，图案
floppy disk 软盘
macrofunction 宏功能
flip flop 触发器
address pointer 地址指针
interface circuitry 接口电路
trench capacitor 沟道电容器
ns *abbr*. nanosecond [ˈnænəuˌsekənd] 纳秒(10^{-9}秒)
MOS *abbr*. Metal-Oxide-Semiconductor 金属氧化物半导体
CMOS *abbr*. Complementary Metal-Oxide-Semiconductor 互补金属氧化物半导体
ECL *abbr*. Emitter Coupled Logic 射极耦合逻辑
UV *abbr*. ultraviolet [ˈʌltrəˈvaiəlit] 紫外线

Notes

1. “铁电”现象是指某些材料受一定强度电场作用时会出现极化或改变原极化方向。
2. 高速缓存用来存储“频繁使用的”或“近期使用的”数据，有助于提高数据读取速度。高速缓存分为两类：(1)位于处理器内的“内部缓存”(或“存储器缓存”)；(2)位于主板上的“外部缓存”(或“磁盘缓存”)。
3. CMOS 集成电路的输出由一个 n 型 MOSFET(即“金属－氧化物－半导体”场效应管)和一个 p 型 MOSFET 串联而成，二者相互“补偿”(complementary)。
4. CAM 是指“内容寻址存储器”。CAM 用来实现高速查找表，但其存储容量要明显小于常规的 DRAM 和 SRAM 存储芯片。
5. 在本句中，whereas 表示转折。此外，whereas 还可以表达“同时”、“鉴于”等涵义。
6. 在芯片制造过程中，要使用“掩模”工艺。当被紫外线照射时，掩模就会在芯片每一层形成不同的电路图案。

Lesson 3 Microprocessors

A microprocessor is a complete computation engine that is fabricated on a single chip. The first microprocessor was the Intel 4004, introduced in 1971. The 4004 was not very powerful - all it could do was to add and subtract, and it could only do that 4

bits at a time. But it was amazing that everything was on one chip. Prior to the 4004, engineers built computers either from collections of chips or from discrete components. The 4004 powered one of the first portable electronic calculators.

The first microprocessor to make it into a home computer was the Intel 8080, a complete 8-bit computer on one chip, introduced in 1974. The first microprocessor to make a real splash in the market was the Intel 8088, introduced in 1979 and incorporated into the IBM PC. The PC market moved from the 8088 to the 80286 to the 80386 to the 80486 to the Pentium to the Pentium II to the Pentium III to the Pentium 4. All of these microprocessors are made by Intel and all of them are improvements on the basic design of the 8088. The Pentium 4 can execute any piece of code that ran on the original 8088, but it does it about 5,000 times faster!

The following table shows the differences between the different processors that Intel has introduced over the years. (See Table 3.1)

Table 3.1 Intel x86 microprocessors

Name	Date	Transistors	Microns[1]	Clock speed	Data width[2]	MIPS[3]
8080	1974	6 000	6	2 MHz	8 bits	0.64
8088	1979	29 000	3	5 MHz	16 bits 8-bit bus	0.33
80286	1982	134 000	1.5	6 MHz	16 bits	1
80386	1985	275 000	1.5	16 MHz	32 bits	5
80486	1989	1 200 000	1	25 MHz	32 bits	20
Pentium	1993	3 100 000	0.8	60 MHz	32 bits 64-bit bus	100
Pentium Ⅱ	1997	7 500 000	0.35	233 MHz	32 bits 64-bit bus	~300
Pentium Ⅲ	1999	9 500 000	0.25	450 MHz	32 bits 64-bit bus	~510
Pentium 4	2000	42 000 000	0.18	1.5 GHz	32 bits 64-bit bus	~1 700

From this table you can see that, in general, there is a relationship between clock speed and MIPS. The maximum clock speed is a function of the manufacturing process and delays within the chip. There is also a relationship between the number of transistors and MIPS. For example, the 8088 clocked at 5 MHz but only executed at 0.33 MIPS (about one instruction per 15 clock cycles). Modern processors can often execute at a rate of two instructions per clock cycle. That improvement is directly related to the number of transistors on the chip.

Inside a Microprocessor

A microprocessor executes a collection of machine instructions that tell the processor what to do. Based on the instructions, a microprocessor does three basic things:

1. Using its ALU (Arithmetic/Logic Unit), a microprocessor can perform mathematical operations like addition, subtraction, multiplication and division. Modern microprocessors contain complete floating point processors that can perform extremely sophisticated operations on large floating point numbers.

2. A microprocessor can move data from one memory location to another.

3. A microprocessor can make decisions and jump to a new set of instructions based on those decisions.

There may be very sophisticated things that a microprocessor does, but those are its three basic activities. The Figure 3.1 shows an extremely simple microprocessor capable of doing those three things:

This microprocessor has an address bus that sends an address to memory , a data bus that can send data to memory or receive data from memory, an RD (read) and WR (write) line to tell the memory whether it wants to set or get the addressed location , a clock line that lets a clock pulse sequence the processor and a reset[4] line that resets the program counter to zero (or whatever) and restarts execution. And let's assume that both the address and data buses are 8 bits wide here.

Here are the components of this simple microprocessor(Figure 3.1):

1. Registers A, B and C are simply latches made out of flip-flops.

2. The address latch is just like registers A, B and C.

3. The program counter is a latch with the extra ability to increment by 1 when told to do so, and also to reset to zero when told to do so.

4. The ALU could be as simple as an 8-bit adder, or it might be able to add, subtract, multiply and divide 8-bit values. Let's assume the latter here.

5. The test register is a special latch that can hold values from comparisons performed in the ALU. An ALU can normally compare two numbers and determine if they are equal, if one is greater than the other, etc. The test register can also normally hold a carry bit from the last stage of the adder. It stores these values in flip-flops and then the instruction decoder can use the values to make decisions.

6. There are six boxes marked "3-State" in the diagram. These are tri-state buffers[5]. A tri-state buffer can pass a 1, a 0 or it can essentially disconnect its output. A tri-state buffer allows multiple outputs to connect to a wire, but only one of them to

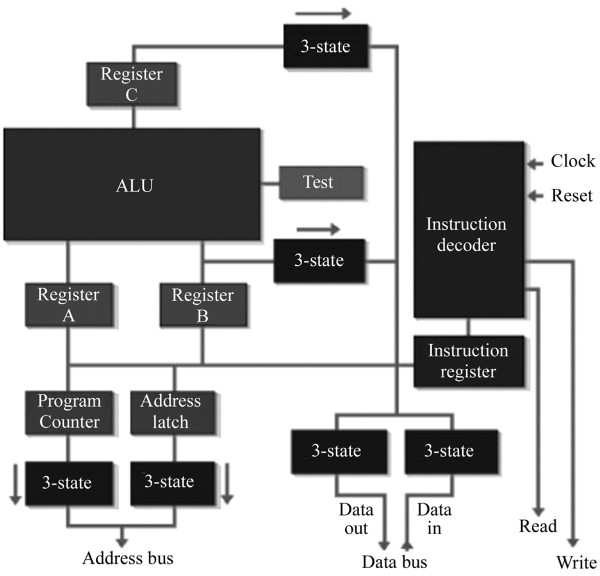


Figure 3.1　A simple microprocessor

actually drive a 1 or a 0 onto the line.

7. The instruction register and instruction decoder are responsible for controlling all of the other components.

Although they are not shown in this diagram, there would be control lines from the instruction decoder that would:

1. Tell the A register to latch the value currently on the data bus
2. Tell the B register to latch the value currently on the data bus
3. Tell the C register to latch the value currently on the data bus
4. Tell the program counter register to latch the value currently on the data bus
5. Tell the address register to latch the value currently on the data bus
6. Tell the instruction register to latch the value currently on the data bus
7. Tell the program counter to increment
8. Tell the program counter to reset to zero
9. Activate any of the six tri-state buffers (six separate lines)
10. Tell the ALU what operation to perform
11. Tell the test register to latch the ALU's test bits
12. Activate the RD line
13. Activate the WR line

Coming into the instruction decoder are the bits from the test register and the clock

line, as well as the bits from the instruction register.

RAM and ROM

The address and data buses, as well as the RD and WR lines connect either to RAM or ROM—generally both. In our sample microprocessor, we have an address bus 8 bits wide and a data bus 8 bits wide. That means that the microprocessor can address (2^8) 256 bytes of memory, and it can read or write 8 bits of the memory at a time. Let's assume that this simple microprocessor has 128 bytes of ROM starting at address 0 and 128 bytes of RAM starting at address 128.

ROM stands for read-only memory. A ROM chip is programmed with a permanent collection of pre-set bytes. The address bus tells the ROM chip which byte to get and place on the data bus. When the RD line changes state, the ROM chip presents the selected byte onto the data bus.

RAM stands for random-access memory. RAM contains bytes of information, and the microprocessor can read or write to those bytes depending on whether the RD or WR line is signaled. One problem with today's RAM chips is that they forget everything once the power goes off. That is why the computer needs ROM.

By the way, nearly all computers contain some amount of ROM (it is possible to create a simple computer that contains no RAM — many microcontrollers do this by placing a handful of RAM bytes on the processor chip itself—but generally impossible to create one that contains no ROM). On a PC, the ROM is called the BIOS (Basic Input/Output System). When the microprocessor starts, it begins executing instructions it finds in the BIOS. The BIOS instructions do things like test the hardware in the machine, and then it goes to the hard disk to fetch the boot sector. This boot sector is another small program, and the BIOS stores it in RAM after reading it off the disk. The microprocessor then begins executing the boot sector's instructions from RAM. The boot sector program will tell the microprocessor to fetch something else from the hard disk into RAM, which the microprocessor then executes, and so on. This is how the microprocessor loads and executes the entire operating system.

Microprocessor Instructions

Even the incredibly simple microprocessor shown here will have a fairly large set of instructions that it can perform. The collection of instructions is implemented as bit patterns, each one of which has a different meaning when loaded into the instruction register. Humans are not particularly good at remembering bit patterns, so a set of

short words are defined to represent the different bit patterns. This collection of words is called the assembly language of the processor. An assembler can translate the words into their bit patterns very easily, and then the output of the assembler is placed in memory for the microprocessor to execute. If you use C language programming, a C compiler will translates the C code into assembly language.

So now the question is, "How do all of these instructions look in ROM?"Each of these assembly language instructions must be represented by a binary number. These numbers are known as opcodes. The instruction decoder needs to turn each of the opcodes into a set of signals that drive the different components inside the microprocessor. Let's take the ADD instruction as an example and look at what it needs to do:

During the first clock cycle, we need to actually load the instruction. Therefore the instruction decoder needs to:

1. activate the tri-state buffer for the program counter
2. activate the RD line
3. activate the data-in tri-state buffer
4. latch the instruction into the instruction register

During the second clock cycle, the ADD instruction is decoded. It needs to do very little:

1. set the operation of the ALU to addition
2. latch the output of the ALU into the C register

During the third clock cycle, the program counter is incremented (in theory this could be overlapped into the second clock cycle).

Every instruction can be broken down as a set of sequenced operations like these that manipulate the components of the microprocessor in the proper order. Some instructions, like this ADD instruction, might take two or three clock cycles. Others might take five or six clock cycles.

Microprocessor Performance

The number of transistors available has a huge effect on the performance of a processor. As seen earlier, a typical instruction in a processor like an 8088 took 15 clock cycles to execute. Because of the design of the multiplier, it took approximately 80 cycles just to do one 16-bit multiplication on the 8088. With more transistors, much more powerful multipliers capable of single-cycle speeds become possible.

More transistors also allow for a technology called pipelining[6]. In a pipelined architecture, instruction execution overlaps. So even though it might take five clock

cycles to execute each instruction, there can be five instructions in various stages of execution simultaneously. That way it looks like one instruction completes every clock cycle.

Many modern processors have multiple instruction decoders, each with its own pipeline. This allows for multiple instruction streams, which means that more than one instruction can complete during each clock cycle. This technique can be quite complex to implement, so it takes lots of transistors.

The trend in processor design has been toward full 32-bit ALUs with fast floating point processors built in and pipelined execution with multiple instruction streams. There has also been a tendency toward special instructions that make certain operations particularly efficient. There has also been the addition of hardware virtual memory support and L1 caching on the processor chip. All of these trends push up the transistor count, leading to the multi-million transistor powerhouses available today. These processors can execute about one billion instructions per second!

New Words

fabricate [ˈfæbrikeit] *vt.* 制作,构造
power [ˈpauə] *n.* 幂,功率 *vt.* 提供动力
splash [splæʃ] *n.* 泼溅
subtraction [səbˈtrækʃn] *n.* 减法
multiplication [ˌmʌltipliˈkeiʃn] *n.* 乘法;繁殖
division [diˈviʒn] *n.* 除法
reset [ˈriːset] *v.* 复位
assume [əˈsjuːm] *vt.* 假定
disconnect [ˌdiskəˈnekt] *v.* 拆开
activate [ˈæktiveit] *vt.* 激活
permanent [ˈpəːmənənt] *adj.* 永久的
signal [ˈsignl] *n.* 信号
overlap [ˈəuvəˈlæp] *v.* 重叠

Phrases & Expressions

make a splash 引起轰动,获得成功
allow for 虑及,酌留

Technical Terms

microprocessor [maikrəʊ'prəʊsesə] *n.* 微处理器
micron ['maikrɔn] *n.* 微米
latch [lætʃ] *n.* 锁存器
buffer ['bʌfə] *n.* 缓冲器
microcontroller [maikrəkən'trəulə] *n.* 微控器
assembler [ə'semblə] *n.* 汇编器
pipelining ['paipˌlainiŋ] *n.* 流水线操作
tri-state buffer 三态缓冲器
boot sector 引导扇区
assembly language 汇编语言
opcode ['ɔpkəud] *abbr.* operating code 操作码
MIPS [mips] *abbr.* Million Instructions Per Second 百万条指令每秒
ALU *abbr.* Arithmetic Logic Unit 算术逻辑单元
BIOS ['baiɔs] *abbr.* Basic Input Output System 基本输入输出系统

Notes

1. micron 即微米(10^{-6}米),记作 μm。micron 用于度量芯片中的最小线宽。micron 值越小,芯片中集成的晶体管数量就越多。
2. 此处的"数据宽度"是指 ALU 的字长。8-bit 的 ALU 可以进行两个 8-bit 数的加、减、乘等运算,32-bit 的 ALU 可以对 32-bit 数进行操作。外部数据总线宽度和 ALU 数据宽度往往相同,但也有例外。如 8088 的 ALU 是 16-bit,其外部数据总线为 8-bit;Pentium 处理器的 ALU 是 32-bit,其外部数据总线为 64-bit。
3. MIPS 是一个粗略表征微处理器性能的指标。因为现代微处理器的种类繁多、结构各不相同,MIPS 指标所能反映的信息很有限。
4. 复位是指"使系统状态处于指定的初始值"。
5. "三态缓冲器"用来控制向总线输出信号。当控制位(control bit)有效时,输入信号通过"三态缓冲器"向总线输出;当控制位无效(inactive)时,"三态缓冲器"的输出端呈现"高阻"(high-impedance,记作 Z)状态,即不向总线输出任何信号。
6. 流水线是一种在前一条指令全部执行完之前就开始取下一条指令、以提高处理速度的技术。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) The microprocessor is the central ______(部件)of the PC. ______ (你做的全部工作)on your computer is performed directly or ______ by the processor. Obviously, it is one of the most important components of the PC, ______(如果不是)the most important. It is also, scientifically, not only one of the most ______(令人惊奇的)parts of the PC, but one of the most amazing ______ (器件)in the world of technology.

The processor ______(扮演重要的角色)in the following important ______(方面)of your computer system:

Performance: The processor is probably the most important single ______(决定性的因素)of system performance in the PC. ______ other components also play a key role in determining performance, the processor's capabilities ______(控制) the maximum performance of a system. The other devices only allow the processor to ______(达到它的全部潜能).

Software Support: Newer, faster processors ______ the use of the latest software. ______ addition, new processors ______ the Pentium with MMX Technology, enable the use of specialized software not usable on earlier machines.

Reliability and Stability: The quality of the processor is one factor that determines ______(系统运行的可靠性). While most processors are very ______(可靠的), some are not. This also depends ______(在某种程度上)on the age of the processor and how much energy it ______(消耗).

Energy Consumption and Cooling: Originally processors consumed relatively little power compared ______ other system devices. Newer processors can consume ______(大量的) power. Power consumption has an impact ______ everything from cooling method selection to overall system reliability.

Motherboard Support: The processor you decide to use in your system will be a ______(主要的)determining factor in what sort of chipset you must use, and hence what motherboard you buy. The motherboard ______(反过来)dictates many facets of your system's capabilities and performance.

(2)Flash memory is a ______(固态的)storage device-everything is ______(电子的). Flash memory provides a ______(非易失的), reliable, low power, low cost, ______(高密度)storage device for programmable ______(代码和数据), ______ it

extremely useful in the ______(嵌入式)marketplace. The most noticeable attribute ______ the FLASH part is its ______ to retain data without the need for power or battery ______(备份). For example, let's say I had my PDA and I was typing in a casual business acquaintance's telephone number. In the middle of my typing, the battery dies. If the information is in RAM, the information will be lost ______ the battery dies. If the information is in FLASH before I lose power, the information will be ______ when I find another battery.

______(从开发者的角度看), it may be important to store data ______ it can be retrieved at a later time, so they use FLASH memory. (However, FLASH memory isn't the only method to accomplish the ______(离线的)storage of data.)

The disadvantage of FLASH memory is ______ it is very difficult to program. Care must be ______ when reading and writing, ______ there are special procedures which need to be performed in order to get the data on the part. ______(因为)the characteristics of the part, it is impossible to write and then immediately overwrite data. In order to write new data to FLASH existing data must be ______ first. The erase process must be done using large blocks of memory. ______ just erasing a portion of the part, you must erase entire sections. This cumbersome process of read/erase/write can get very complicated when there are ______(重复的)write operations being performed.

Another aspect of flash memory ______(必须要考虑的)is its life expectancy: Continuous erasing can ______(对……产生破坏性的影响)a FLASH part. ______(对任一给定的闪存), there is a limit ______ the total number of erase operations that may be performed on a particular erase sector before it becomes unreliable ______ damaged.

Despite the complications, there are many advantages to FLASH memory, so why don't we just use FLASH memory for everything? The cost per megabyte for a hard disk is cheaper, and the capacity is much greater.

2. Translate the following passages into Chinese or English.

1) Many of the products and services in modern society are based upon the work of electrical engineers and computer scientists. The tremendous reduction over the last decade in the cost of digital electronic devices has led to an explosive growth in the use of computers and computation. At the same time, our increased understanding of computer science has made possible the development of new software systems of increased power, sophistication, and flexibility.

2) In November 1971, Intel introduced the world's first commercial microprocessor, the 4004, invented by three Intel engineers. Primitive by today's standards, it con-

tained a mere 2 300 transistors and performed about 60,000 calculations in a second. Today, the microprocessor is the most complex mass-produced product ever, with millions of transistors performing hundreds of millions of calculations each second.

3) Microprocessors are the brains of your personal computer. Here on this tiny silicon chip are millions of switches and pathways that help your computer make important decisions and perform helpful tasks. And microprocessors don't just think for computers-you might find a processor in many other everyday items like your telephone or car.

4) Memory is the component of a computer (and an embedded system) that is used to store and retain information. There is always a need to store information in an embedded application, so there will always be some type of memory in the design. Many types of memory are available, thus there are many choices for embedded customers.

5) Memory can be split into two main categories: volatile and nonvolatile. Volatile memory loses its content when the system is powered off. Most types of RAM fall into this category. Nonvolatile memory does not lose its data when the system or device is turned off.

6) 现代电子系统的数字化程度越来越高,复杂程度也极高。要是没有超大规模芯片技术,这是无法想象的。现代电子系统太复杂了,以致其设计原理很像大型软件系统设计。计算机科学和电子系统设计所需的背景知识存在许多相同之处,而现代电子系统——这个不断扩展的工程领域——也离不开计算机辅助设计。

7)“芯片”就是一小片嵌入了集成电路的半导体材料。芯片的典型尺寸不超过一个平方英寸,却能包含几百万个晶体管。芯片的类型多种多样。例如,CPU 芯片内含完整的处理单元,而存储芯片包含的是空白存储单元。

8) 微处理器就是一片包含了 CPU 的硅片。用来区分微处理器的基本特征有三个:指令集、带宽和时钟速度。

9) 不用硬盘而采用闪存的理由有好几个:闪存的存取速度更快、不产生噪声、重量更轻、尺寸更小,而且没有可以活动的部件。

10) 在计算机系统中,存储器是用于程序运行和数据存储的部件。随机存取存储器芯片构成了计算机的主存储器。可用存储器的容量决定了运行程序的大小和一次能否运行多个程序。主存储器只是暂时储存数据。一旦关闭计算机,数据就丢失了。它不同于内部只读存储器和外部存储器,后者的数据保存时间更为长久。只读存储器中包含着计算机的核心程序。总之,任何以机器可读方式存储数据的设备都是存储器。

Reading Materials

Passage 1 The Digital Age

The Analog Age

Up until the middle of the 20th century, the technology designed by engineers was primarily analog; more specifically, the devices and systems that engineers created relied primarily upon physical forces and matter for their basic operation rather than some abstract quantity, such as numbers.

For example, analog audio records introduced in the first part of the 20th century use the physical bumps and indentations in the grooves on vinyl discs (albums) to store audio data. The stylus at the end of the tone arm of a turntable rides in these grooves and vibrates nearly identically to the original sound waves of the audio. This mechanical motion is then converted to an identical electrical version of the audio that is subsequently amplified and played through speakers.

The entire process of re-creating audio from bumps and indentations is purely physical. Never is the sound waveform converted to numbers to be stored or manipulated; in other words, the system is analog.

Analog systems like turntables and albums are quite functional. However, like all designs, they suffer from a number of shortcomings:

- Analog systems can be large.
- Analog systems can consume lots of energy.
- Analog systems are not easily modified to solve new or different tasks.
- Analog systems are prone to breakdowns due to their physical operation.

Building Blocks for Analog Designs

Early electronic technology was built using an important analog device called a vacuum tube. Such tubes were used to control the electrical current and voltage in systems such as radios, radar, and very early computers. Unfortunately, like light bulbs, these vacuum tubes got very hot, and burned out regularly. Your older family members might remember how often TVs used to break down due to vacuum tubes burning out.

The Birth of the Digital Age

During the middle of the 20th century, mathematicians and engineers discovered a process for converting most physical quantities found in the world (such as sound waves, light intensity, forces, voltage, current, or charge) to numbers or digits. This discovery should not be surprising, since scientists had been using mathematics to describe the physical world for centuries. This remarkable, yet simple, discovery was the mathematical foundation that gave birth to the digital age.

There are many advantages to "digitizing" analog quantities. For example:

- Numbers are much less sensitive to physical problems caused by the physical nature of the device used to store or manipulate them.
- Numbers are easier to store than an equivalent physical "amount" of something.
- Numbers can be moved through space, using electronic, optical, or acoustic means.

To illustrate these advantages, let's revisit the re-creation of audio systems as discussed earlier in this section. Unlike analog systems, today the sounds of most audio are converted to numbers at the recording studio and stored on a compact disc (CD) or DVD. A CD player simply reads the numbers from the CD and then converts these numbers back to the original audio. If you have ever compared the quality of audio between an average turntable and an average CD player, you should have little doubt that digital technology is significantly better than the earlier analog technology.

Also, can you imagine trying to jog or walk with a turntable strapped to your waist or inside your backpack? Table 1 lists some analog devices and the corresponding digital devices. Which do you prefer to use?

Table 1

Analog Devices	Digital Devices
LPs	CDs
Film cameras	Digital cameras
Dial watches	Digital watches
Standard TV	HDTV
VHS camcorders	Digital camcorders

Still a Long Way To Go

Unfortunately, there was a major problem in building new digital devices when they

were first conceived. Engineers just didn't have the right parts to build new digital systems. Not to be deterred, engineers working during the first half of the 20th century tried the smart and reasonable thing: They attempted to use readily available vacuum tubes as basic digital building blocks. Following this approach, in 1945, engineers successfully produced the first digital computer, called the ENIAC. It was built out of more than 17,000 vacuum tubes, weighed 30 tons, and filled a 30-by- 50-foot room. Just think of the heat produced by 17,000 light bulbs all burning in the same room!

While primitive by today's standards, the ENIAC was a major advance in engineering and technology. Never before in human history could we do math so fast and so accurately. While the ENIAC opened up new digital horizons for society, these first computers were so large and so expensive that only governments and the largest of companies could ever hope to own or even use one.

The Transistor Replaced the Vacuum Tube

What the digital age needed was a truly digital component that could replace the vacuum tube. It would have to run fast, use much less power than the vacuum tube, and, most importantly, be small and inexpensive.

Fortunately, in 1947 engineers at AT&T's Bell Laboratories developed that component, called the transistor. Its creation changed the world forever. Bill Shockley, Walter Brattain, and John Bardeen won the Nobel Prize in 1956 for their joint discovery and development of the transistor, which, within a few decades, had completely replaced the vacuum tube in nearly every piece of technology.

Now, engineers could unleash their imaginations to create smaller, portable devices that could run on the relatively small amounts of energy contained in batteries and were rugged in normal use.

For this reason, many people believe that the transistor is the most important invention of the 20th century. Just look around you today to see the nearly infinite array of small gadgets and pieces of technology built from transistors.

The Integrated Circuit (IC)

As engineers designed devices for more complex tasks, such as in robotics or medicine, the resulting systems required ever more transistors. This push for more transistors made the devices large and hard to wire together.

The next critical step forward into the digital age was the ability to put many transistors onto a single small part that could be used for these increasingly complex

tasks. Jack Kilby accomplished this remarkable feat in 1958 at Texas Instruments when he invented the integrated circuit, or IC.

For this discovery, Kilby was awarded the 2000 Nobel Prize in physics. This groundbreaking invention was coined the "integrated circuit" because it cleverly integrated many parts, typically transistors, into a single small package.

With the invention of the IC, engineers were able to undertake more complicated designs, because they now had modern digital parts that could do significantly more complicated math on the newly digitized version of the real analog world. Interestingly, the integrated circuit has become so pervasive in devices from computers to anti-lock brakes that it is difficult to find individual transistors in modern technology today—they are now all part of integrated circuits.

Questions:

1) What shortcomings does the analog technology have?

2) Which one do you prefer a film camera or a digital camera? Why?

3) Could you say something about the vacuum tube?

4) Why is the transistor regarded as the most important invention in the 20th century?

5) What does the term ***IC*** stand for?

Passage 2 Flash memory

Electronic memory comes in a variety of forms to serve a variety of purposes. Flash memory is used for easy and fast information storage in such devices as digital cameras and home video game consoles. It is used more as a hard drive than as RAM. In fact, Flash memory is considered a solid state storage device. Solid state means that there are no moving parts - everything is electronic instead of mechanical.

Here are a few examples of Flash memory:

- Your computer's BIOS chip
- CompactFlash (most often found in digital cameras)
- SmartMedia (most often found in digital cameras)
- Memory Stick (most often found in digital cameras)
- PCMCIA Type Ⅰ and Type Ⅱ memory cards (used as solid-state disks in laptops)

- Memory cards for video game consoles

Flash memory is a type of EEPROM chip. It has a grid of columns and rows with a cell that has two transistors at each intersection. The two transistors are separated from each other by a thin oxide layer. One of the transistors is known as a floating gate, and the other one is the control gate. The floating gate's only link to the row, or wordline, is through the control gate. As long as this link is in place, the cell has a value of 1. To change the value to a 0 requires a curious process called Fowler-Nordheim tunneling.

Tunneling is used to alter the placement of electrons in the floating gate. An electrical charge, usually 10 to 13 volts, is applied to the floating gate. The charge comes from the column, or bitline, enters the floating gate and drains to a ground.

This charge causes the floating-gate transistor to act like an electron gun. The excited electrons are pushed through and trapped on other side of the thin oxide layer, giving it a negative charge. These negatively charged electrons act as a barrier between the control gate and the floating gate. A special device called a cell sensor monitors the level of the charge passing through the floating gate. If the flow through the gate is greater than 50 percent of the charge, it has a value of 1. When the charge passing through drops below the 50-percent threshold, the value changes to 0. A blank EEPROM has all of the gates fully open, giving each cell a value of 1.

The electrons in the cells of a Flash-memory chip can be returned to normal ("1") by the application of an electric field, a higher-voltage charge. Flash memory uses in-circuit wiring to apply the electric field either to the entire chip or to predetermined sections known as blocks. This erases the targeted area of the chip, which can then be rewritten. Flash memory works much faster than traditional EEPROMs because instead of erasing one byte at a time, it erases a block or the entire chip, and then rewrites it.

You may think that your car radio has Flash memory, since you are able to program the presets and the radio remembers them. But it is actually using Flash RAM. The difference is that Flash RAM has to have some power to maintain its contents, while Flash memory will maintain its data without any external source of power. Even though you have turned the power off, the car radio is pulling a tiny amount of current to preserve the data in the Flash RAM. That is why the radio will lose its presets if your car battery dies or the wires are disconnected.

Although standards are flourishing, there are many Flash-memory products that are completely proprietary in nature, such as the memory cards in video game systems. But it is good to know that as electronic components become increasingly interchangeable and learn to communicate with each other (by way of technologies such as Bluetooth),

standardized removable memory will allow you to keep your world close at hand.

(Notes: PCMCIA- The Personal Computer Memory Card International Association (PCMCIA) was formed by several card manufacturers in the late 1980s. The PCMCIA standard allows for the PC Card to be used with many computer types regardless of the microprocessor type. Not only can PC Cards be used with many types of computers, they are also suitable for use in other digital applications such as test equipment, digital imaging equipment and industrial controllers. PC Cards are now used in many applications, including RAM memory, pre-programmed ROM cards, modems, sound cards, floppy disk controllers, hard drives, CD ROM and SCSI controllers, Global Positioning System (GPS) cards, data acquisition, LAN cards, pagers, etc..)

Questions:

1) What does the author mean by saying "*It is used more as a hard drive than as RAM*"?

2) What does the term ***solid state device*** refer to?

3) Which is the normal state in a Flash-memory chip, "1" or "0"?

4) How does Flash memory manage to work faster than traditional EEPROMs?

5) What's the difference between a Flash memory and a Flash RAM?

Passage 3 Microcontrollers

Microcontrollers are hidden inside a surprising number of products these days. If your microwave oven has an LED or LCD screen and a keypad, it contains a microcontroller. All modern automobiles contain at least one microcontroller, and can have as many as six or seven: The engine is controlled by a microcontroller, as are the anti-lock brakes, the cruise control and so on. Any device that has a remote control almost certainly contains a microcontroller: TVs, VCRs and high-end stereo systems all fall into this category. Nice SLR and digital cameras, cell phones, camcorders, answering machines, laser printers, telephones (the ones with caller ID, 20-number memory, etc.), pagers, and feature-laden refrigerators, dishwashers, washers and dryers (the ones with displays and keypads)… Basically, any product or device that interacts with its user has a microcontroller buried inside.

What is a Microcontroller? A microcontroller is a computer. All computers - whether a personal desktop computer or a large mainframe computer or a microcontroller - have several things in common:

All computers have a CPU that executes programs.

The CPU loads the program from somewhere (e.g. hard disks).

The computer has some RAM where it can store variables.

And the computer has some input and output devices so it can talk to people.

The desktop computer you are using is a general-purpose computer that can run any of thousands of programs. Microcontrollers are special purpose computers, which do one thing well. If a computer matches a majority of these characteristics, then you can call it a microcontroller:

Microcontrollers are "embedded" inside some other device (often a consumer product) so that they can control the features or actions of the product.

Microcontrollers are dedicated to one task and run one specific program. The program is stored in ROM (Read-Only Memory) and generally does not change.

Microcontrollers are often low-power devices. A desktop computer is almost always plugged into a wall socket and might consume 50 watts of electricity. A battery-operated microcontroller might consume 50 milliwatts.

A microcontroller has a dedicated input device and often (but not always) has a small LED or LCD display for output. For example, the microcontroller inside a TV takes input from the remote control and displays output on the TV screen. The controller controls the channel selector, the speaker system and certain adjustments on the picture tube electronics such as tint and brightness.

A microcontroller is often small and low cost. The components are chosen to minimize size and to be as inexpensive as possible.

A microcontroller is often, but not always, ruggedized in some way.

The microcontroller controlling a car's engine, for example, has to work in temperature extremes that a normal computer generally cannot handle. A car's microcontroller in Alaska has to work fine in −30 degree F (−34 ℃) weather, while the same microcontroller in Nevada might be operating at 120 degrees F (49 ℃).

The actual processor used to implement a microcontroller can vary widely. In many products, such as microwave ovens, the demand on the CPU is fairly low and price is an important consideration. In these cases, manufacturers turn to dedicated microcontroller chips - chips that were originally designed to be low-cost, small, low-power, embedded CPUs. The Motorola 6811 and Intel 8051 are both good examples of such chips. There is also a line of popular controllers called "PIC microcontrollers" created by a company called Microchip. By today's standards, these CPUs are incredibly minimalistic; but they are extremely inexpensive when purchased in large quantities and can often meet the needs of a device's designer with just one chip.

Questions：

1）What characteristics do all computers share?

2）How is a microcontroller defined?

3）What functions does the microcontroller inside a TV perform?

4）Why does a microcontroller often require being ruggedized in some way?

5）What kinds of products are dedicated microcontroller chips often applied to?

Unit 2

Electronic Circuits

Lesson 4 Operational Amplifiers

In 1934 Harry Black commuted from his home in New York City to work at Bell Labs in New Jersey by way of a railroad/ferry. The ferry ride relaxed Harry enabling him to do some conceptual thinking. Harry had a tough problem to solve; when phone lines were extended long distances, they needed amplifiers, and undependable amplifiers limited phone service. First, initial tolerances on the gain were poor, but that problem was quickly solved with an adjustment. Second, even when an amplifier was adjusted correctly at the factory, the gain drifted so much during field operation that the volume was too low or the incoming speech was distorted.

Many attempts had been made to make a stable amplifier, but temperature changes and power supply voltage extremes experienced on phone lines caused uncontrollable gain drift. Passive components had much better drift characteristics than active components had, thus if an amplifier's gain could be made dependent on passive components, the problem would be solved. During one of his ferry trips, Harry's fertile brain conceived a novel solution for the amplifier problem, and he documented the solution while riding on the ferry.

The solution was to first build an amplifier that had more gain than the application required. Then some of the amplifier output signal was fed back to the input in a manner that makes the circuit gain (circuit is the amplifier and feedback components) dependent on the feedback circuit rather than the amplifier gain. Now the circuit gain is dependent on the passive feedback components rather than the active amplifier. This is called negative feedback, and it is the underlying operating principle for all modern day op amps. Harry had documented the first intentional feedback circuit during a ferry ride. I am sure unintentional feedback circuits had been built prior to that time, but the designers ignored the effect!

I can hear the squeals of anguish coming from the managers and amplifier designers. I imagine that they said something like this, "it is hard enough to achieve 30-kHz gainbandwidth (GBW), and now this fool wants me to design an amplifier with 3-MHz GBW. But, he is still going to get a circuit gain GBW of 30 kHz." Well, time has proven Harry right, but there is a minor problem that Harry didn't discuss in detail, and that is the oscillation problem. It seems that circuits designed with large open loop

gains sometimes oscillate when the loop is closed. A lot of people investigated the instability effect, and it was pretty well understood in the 1940s, but solving stability problems involved long, tedious, and intricate calculations. Years passed without anybody making the problem solution simpler or more understandable.

In 1945 H. W. Bode presented a system for analyzing the stability of feedback systems by using graphical methods. Until this time, feedback analysis was done by multiplication and division, so calculation of transfer functions was a time consuming and laborious task. Remember, engineers did not have calculators or computers until the '70s. Bode presented a log technique that transformed the intensely mathematical process of calculating a feedback system's stability into graphical analysis that was simple and perceptive. Feedback system design was still complicated, but it no longer was an art dominated by a few electrical engineers kept in a small dark room. Any electrical engineer could use Bode's methods to find the stability of a feedback circuit, so the application of feedback to machines began to grow. There really wasn't much call for electronic feedback design until computers and transducers become of age.

The first real-time computer was the analog computer! This computer used preprogrammed equations and input data to calculate control actions. The programming was hard wired with a series of circuits that performed math operations on the data, and the hard wiring limitation eventually caused the declining popularity of the analog computer. The heart of the analog computer was a device called an operational amplifier because it could be configured to perform many mathematical operations such as multiplication, addition, subtraction, division, integration, and differentiation on the input signals. The name was shortened to the familiar op amp, as we have come to know and love them. The op amp used an amplifier with a large open loop gain, and when the loop was closed, the amplifier performed the mathematical operations dictated by the external passive components. This amplifier was very large because it was built with vacuum tubes and it required a high-voltage power supply, but it was the heart of the analog computer, thus its large size and huge power requirements were accepted. Many early op amps were designed for analog computers, and it was soon found out that op amps had other uses and were very handy to have around the physics lab.

At this time general-purpose analog computers were found in universities and large company laboratories because they were critical to the research work done there. There was a parallel requirement for transducer signal conditioning in lab experiments, and op amps found their way into signal conditioning applications. As the signal conditioning applications expanded, the demand for op amps grew beyond the analog computer

requirements, and even when the analog computers lost favor to digital computers, the op amp survived because of its importance in universal analog applications. Eventually digital computers replaced the analog computers, but the demand for op amps increased as measurement applications increased.

The first signal conditioning op amps were constructed with vacuum tubes prior to the introduction of transistors, so they were large and bulky. During the '50s, miniature vacuum tubes that worked from lower voltage power supplies enabled the manufacture of op amps that shrunk to the size of a brick used in house construction, so the op amp modules were nicknamed bricks. Vacuum tube size and component size decreased until an op amp was shrunk to the size of a single octal vacuum tube. Transistors were commercially developed in the '60s, and they further reduced op amp size to several cubic inches. Most of these early op amps were made for specific applications, so they were not necessarily general purpose. The early op amps served a specific purpose, but each manufacturer had different specifications and packages; hence, there was little second sourcing among the early op amps.

ICs were developed during the late 1950s and early 1960s, but it wasn't till the middle 1960s that Fairchild released the μA709. This was the first commercially successful IC op amp. The μA709 had its share of problems, but any competent analog engineer could use it, and it served in many different analog applications. The major drawback of the μA709 was stability; it required external compensation and a competent analog engineer to apply it. Also, the μA709 was quite sensitive because it had a habit of self-destruction under any adverse condition. The μA741 followed the μA709, and it is an internally compensated op amp that does not require external compensation if operated under data sheet conditions. There has been a never-ending series of new op amps released each year since then, and their performance and reliability has improved to the point where present day op amps can be used for analog applications by anybody [1].

The IC op amp is here to stay; the latest generation op amps cover the frequency spectrum from 5-kHz GBW to beyond 1-GHz GBW. The supply voltage ranges from guaranteed operation at 0.9 V to absolute maximum voltage ratings of 1000 V. The input current and input offset voltage has fallen so low that customers have problems verifying the specifications during incoming inspection. The op amp has truly become the universal analog IC because it performs all analog tasks. It can function as a line driver [2], comparator (one bit A/D), amplifier, level shifter, oscillator, filter, signal conditioner, actuator driver, current source, voltage source, and etc.. The designer's

problem is how to rapidly select the correct circuit/op amp combination and then, how to calculate the passive component values that yield the desired transfer function in the circuit.

The op amp will continue to be a vital component of analog design because it is such a fundamental component. Each generation of electronic equipment integrates more functions on silicon and takes more of the analog circuitry inside the IC. As digital applications increase, analog applications also increase because the predominant supply of data and interface applications are in the real world, and the real world is an analog world. Thus, each new generation of electronic equipment creates requirements for new analog circuits; hence, new generations of op amps are required to fulfill these requirements. Analog design, and op amp design, is a fundamental skill that will be required far into the future.

New Words

commute [kə'mjuːt] *v.* 通勤

undependable [ˌʌndi'pendəbl] *adj.* 不可靠的

adjustment [ə'dʒʌstmənt] *n.* 调整，调节

drift [drift] *n.* 漂移

underlying [ˌʌndə'laiiŋ] *adj.* 根本的，潜在的

squeal [skwiːl] *v.* 长声尖叫

anguish ['æŋgwiʃ] *n.* 痛苦，苦恼

graphical ['græfikəl] *adj.* 图形的

laborious [lə'bɔːriəs] *adj.* 艰苦的，费力的

perceptive [pə'septiv] *adj.* 有知觉的，有理解力的

handy ['hændi] *adj.* 手边的，容易取得的

bulky ['bʌlki] *adj.* 容量大的，体积大的

octal ['ɔktl] *adj.* 八管脚的，八进制的

cubic ['kjuːbik] *adj.* 立方体的，立方的

adverse ['ædvəːs] *adj.* 不利的，相反的

drawback ['drɔːˌbæk] *n.* 缺点，障碍

sourcing [sɔːsiŋ] *n.* 供货

rating ['reitiŋ] *n.* 等级，级别

predominant [pri'dɔminənt] *adj.* 卓越的，支配的，主要的

Phrases & Expressions

prior to 先于,在……之前
in detail 详细地
of age 成熟;发达;充分发展

Technical Terms

stability [stə'biliti] *n.* 稳定性
conditioning [kən'diʃniŋ] *n.* 调节,调整
transducer [trænz'djuːsə] *n.* 传感器,变换器
self-destruction ['selfdis'trʌkʃn] *n.* 自毁
passive ['pæsiv] *adj.* 无源的
comparator ['kɔmpəreitə] *n.* 比较器
actuator ['æktjueitə] *n.* 致动器,执行器
oscillator ['ɔsileitə] *n.* 振荡器
transfer function 传输函数
data sheet 数据手册
incoming inspection 入厂检查;输入检验
line driver 线路驱动器
level shifter 电平移动器
signal conditioner 信号调节器
current source 电流源
voltage source 电压源
GBW *abbr.* Gain × Bandwidth 增益带宽积

Notes

1. 此句可译为:“此后,新型运算放大器就不断出现。如今,运算放大器的功能和可靠性已经提高到了人人都能在模拟应用中使用它的程度。”to the point where 结构表示“达到……程度”。
2. 线路驱动器可以是用于驱动多个逻辑负载、低阻抗总线 I/O 或远距离传送信号的功率放大器,也可以是用于串/并转换、控制字符插入或数据缓冲的硬件单元。

Lesson 5 Low-pass Filters

First-Order Filters

An integrator(Figure 5.1a) is the simplest filter mathematically, and it forms the building block for most modern integrated filters. Consider what we know intuitively about an integrator. If you apply a DC signal at the input (i.e., zero frequency), the output will describe a linear ramp that grows in amplitude until limited by the power supplies. Ignoring that limitation, the response of an integrator at zero frequency is infinite, which means that it has a pole at zero frequency. (A pole exists at any frequency for which the transfer function's value becomes infinite.)

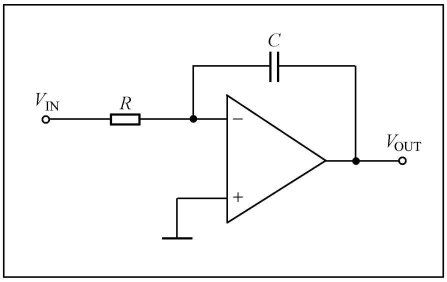


Figure 5.1a A simple RC integrator

We also know that the integrator's gain diminishes with increasing frequency and that at high frequencies the output voltage becomes virtually zero. Gain is inversely proportional to frequency, so it has a slope of -1 when plotted on log/log coordinates (i.e., −20 dB/decade on a Bode plot, Figure 5.1b).

You can easily derive the transfer function as

$$\frac{V_{OUT}}{V_{IN}} = \frac{X_C}{R} = \frac{1/sC}{R} = \frac{\omega_0}{s}$$

where s is the complex-frequency variable $\sigma + j\omega$ and ω_0 is $1/RC$. If we think of s as frequency, this formula confirms the intuitive feeling that gain is inversely proportional to frequency.

The next most complex filter is the simple low-pass RC type (Figure 5.2a). Its characteristic (transfer function) is

$$\frac{V_{OUT}}{V_{IN}} = \frac{1/sC}{R + 1/sC} = \frac{1}{1 + sCR} = \frac{\omega_0}{s + \omega_0}$$

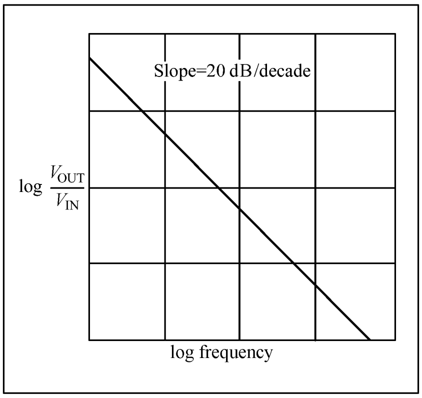


Figure 5.1b A Bode plot of a simple integrator

When $s = \omega_0$, the function reduces to ω_0/ω_0, i.e., 1. When s tends to infinity, the function tends to zero, so this is a low-pass filter. When $s = -\omega_0$, the denominator is zero and the function's value is infinite, indicating a pole in the complex frequency plane. The magnitude of the transfer function is plotted against s in Figure 5.2b, where the real component of s (σ) is toward us and the positive imaginary part (jω) is toward the right. The pole at $-\omega_0$ is evident. Amplitude is shown logarithmically to emphasize the function's form. For both the integrator and the RC low-pass filter, frequency response tends to zero at infinite frequency; that is, there is a zero at $s=\infty$. This single zero surrounds the complex plane.

But how does the complex function in s relate to the circuit's response to actual frequencies? When analyzing the response of a circuit to AC signals, we use the expression jωL for impedance of an inductor and $1/\mathrm{j}\omega C$ for that of a capacitor. When analyzing transient response using Laplace transforms [1], we use sL and $1/sC$ for the impedance of these elements. The similarity is apparent immediately. The jω in AC analysis is in fact the imaginary part of s, which, as mentioned earlier, is composed of a real part σ and an imaginary part jω.

If we replace σ by jω in any equation so far, we have the circuit's response to an angular frequency. In the complex plot in Figure 5.2b, $\sigma = 0$ and hence s = jw along the positive j axis. Thus, the function's value along this axis is the frequency response of the filter. We have sliced the function along the jω axis and emphasized the RC low-pass filter's frequency-response curve by adding a heavy line for function values along the positive j axis. The more familiar Bode plot (Figure 5.2c) looks different in form only because the frequency is expressed logarithmically.

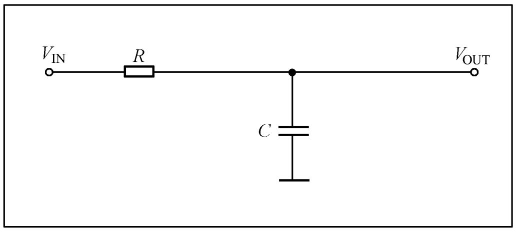


Figure 5.2a A simple RC low-pass filter

While the complex frequency's imaginary part ($j\omega$) helps describe a response to AC signals, the real part (σ) helps describe a circuit's transient response. Looking at Figure 5.2b, we can therefore say something about the RC low-pass filter's response as compared to that of the integrator. The low-pass filter's transient response is more stable, because its pole is in the negative-real half of the complex plane. That is, the low-pass filter makes a decaying-exponential response to a step-function input; the integrator makes an infinite response. For the low-pass filter, pole positions further down the $-\sigma$ axis mean a higher ω_0, a shorter time constant, and therefore a quicker transient response. Conversely, a pole closer to the j axis causes a longer transient response.

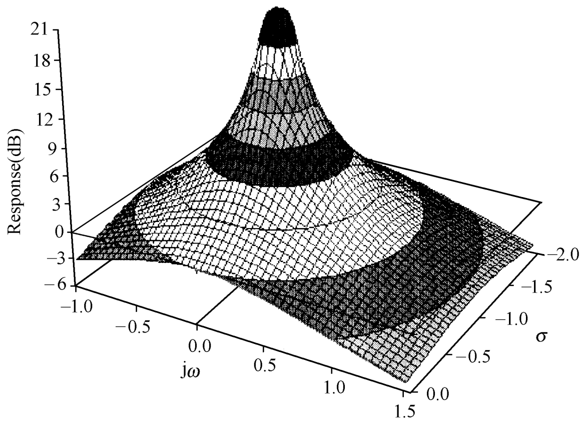


Figure 5.2b The complex function of an RC low-pass filter

So far, we have related the mathematical transfer functions of some simple circuits to their associated poles and zeroes in the complex-frequency plane. From these functions, we have derived the circuit's frequency response (and hence its Bode plot) and also its transient response. Because both the integrator and the RC filter have only one s in the denominator of their transfer functions, they each have only one pole. That is, they are first-order filters.

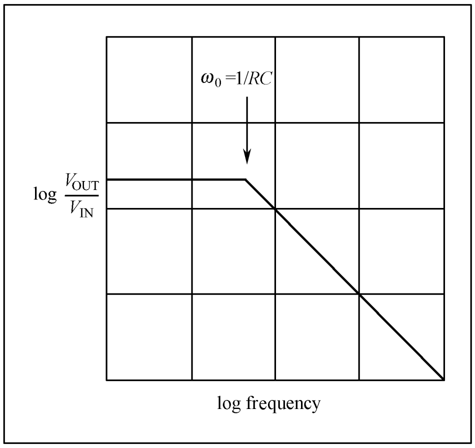


Figure 5.2c A Bode plot of a low-pass filter

However, as we can see from Figure 5.1b, the first-order filter does not provide a very selective frequency response. To tailor a filter more closely to our needs, we must move on to higher orders. From now on, we will describe the transfer function using $f(s)$ rather than the cumbersome V_{OUT}/V_{IN}.

Second-Order Low-Pass Filters

A second-order filter has s^2 in the denominator and two poles in the complex plane. You can obtain such a response by using inductance and capacitance in a passive circuit or by creating an active circuit of resistors, capacitors, and amplifiers. Consider the passive RLC filter in Figure 5.3a, for instance. We can show that its transfer function has the form

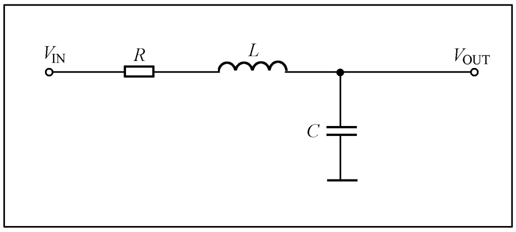


Figure 5.3a An RLC low-pass filter

$$f(s) = \frac{X_C}{R + X_L + X_C} = \frac{1/sC}{R + sL + 1/sC} = \frac{1}{LCs^2 + RCs + 1}$$

and if we define $\omega_0^2 = 1/LC$ and $Q = \omega_0 L/R$, then

$$f(s) = \frac{\omega_0^2}{s^2 + s\omega_0/Q + \omega_0^2}$$

where ω_0 is the filter's characteristic frequency and Q is the quality factor(lower R means higher Q).

The poles occur at s values for which the denominator becomes zero; that is, when $s^2 + s\,\omega_0/Q + \omega_0^2 = 0$. We can solve this equation by remembering that the roots of $ax^2 + bx + c = 0$are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

In this case, $a = 1$, $b = \omega_0/Q$, and $c = \omega_0^2$. The term $(b^2 - 4ac)$ equals $\omega_0^2(1/Q^2 - 4)$, so if Q is less than 0.5 then both roots are real and lie on the negative-real axis. The circuit's behavior is much like that of two first order RC filters in cascade. This case isn't very interesting, so we'll consider only the case where $Q > 0.5$, which means $(b^2 - 4ac)$is negative and the roots are complex.

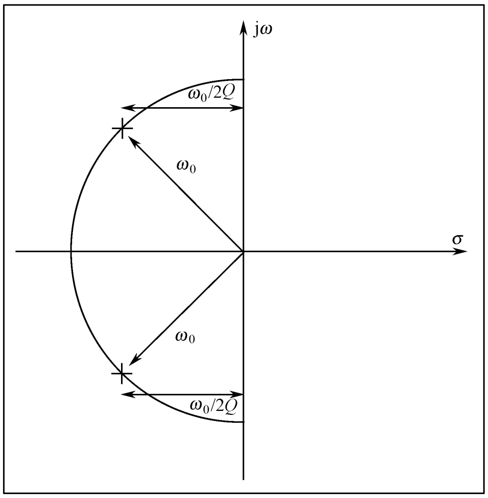


Figure 5.3b A pole-zero diagram of an RLC low-pass filter

The real part is therefore $-b^2 a$, which is $-\omega_0/2Q$, and common to both roots. The roots' imaginary parts will be equal and opposite in signs. Calculating the position of the roots in the complex plane, we find that they lie at a distance of x_0 from the origin, as shown in Figure 5.3b.

Varying ω_0 changes the poles' distance from the origin. Decreasing the Q moves the poles toward each other, whereas increasing the Q moves the poles in a semicircle away from each other and toward the jω axis. When $Q = 0.5$, the poles meet at $-\omega_0$ on the

negative-real axis. In this case, the corresponding circuit is equivalent to two cascaded first-order filters.

Now let's examine the second-order function's frequency response and see how it varies with Q. As before, Figure 5.4a shows the function as a curved surface, depicted in the three-dimensional space formed by the complex plane and a vertical magnitude vector. $Q = 0.707$, and you can see immediately that the response is a low-pass filter.

The effect of increasing the Q is to move the poles in a circular path toward the $j\omega$ axis. Figure 5.4b shows the case where $Q = 2$. Because the poles are closer to the $j\omega$ axis, they have a greater effect on the frequency response, causing a peak at the high end of the passband.

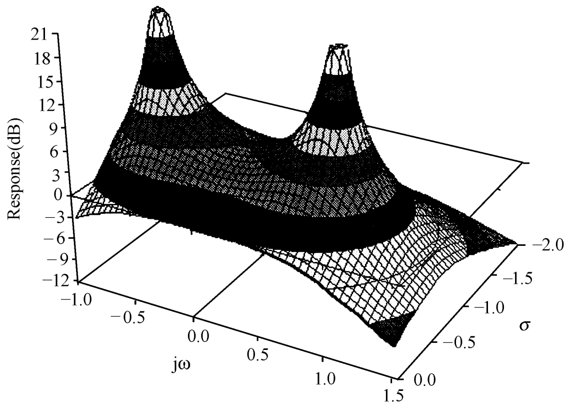


Figure 5.4a The complex function of a second-order low-pass filter ($Q = 0.707$)

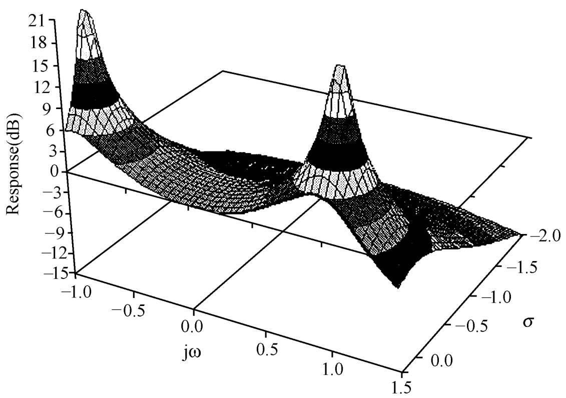


Figure 5.4b The complex function of a second-order low-pass filter ($Q = 2$)

There is also an effect on the filter's transient response. Because the poles'

negative-real part is smaller, an input step function will cause ringing at the filter output. Lower values of Q result in less ringing, because the damping is greater. On the other hand, if Q becomes infinite, the poles reach the $j\omega$ axis, causing an infinite frequency response (instability and continuous oscillation) at $\omega=\omega_0$. In the RLC circuit in Figure 5.3a, this condition would be impossible unless $R=0$. For filters that contain amplifiers, however, the condition is possible and must be considered in the design process.

A second-order filter provides the variables ω_0 and Q, which allow us to place poles wherever we want in the complex plane. These poles must, however, occur as complex-conjugate pairs, in which the real parts are equal and the imaginary parts have opposite signs. This flexibility in pole placement is a powerful tool and one that makes the second-order stage a useful component in many switched-capacitor filters. As in the first-order case, the second-order low-pass transfer function tends to zero as frequency tends to infinity. The second-order function decreases twice as fast, however, because of the s^2 factor in the denominator. The result is a double zero at infinity.

New Words

intuitively [in'tjuitivli] *adv.* 直觉地,直观地
linear ['liniə] *adj.* 线性的
ramp [ræmp] *n.* 斜坡，坡道
inversely ['in'vəːs] *adv.* 相反地,反向地
proportional [prə'pɔːʃnl] *adj.* 成比例的,相称的,均衡的
slope [sləup] *n.* 斜坡，斜面
coordinate [kəu'ɔːdinit] *n.* 坐标
logarithm ['lɔgəriðəm] *n.* 对数
formula ['fɔːmjulə] *n.* 公式,配方,规则
denominator [di'nɔmineitə] *n.* 分母
selective [si'lektiv] *adj.* 选择性的
cumbersome ['kʌmbəsəm] *adj.* 麻烦的,笨重的
origin ['ɔridʒin] *n.* 原点
cascade [kæs'keid] *n.* 级联
semicircle ['semi'səːkl] *n.* 半圆形
transient ['trænziənt] *adj.* 短暂的，瞬时的 *n.* 瞬时现象
exponential [ˌekspəu'nenʃəl] *adj.* 指数的

conjugate [ˈkɔndʒugit] *adj.* 共轭的

Phrases & Expressions

(be) proportional to 与……成比例
(be) equivalent to 相当于……,等价于……
in cascade 级联
have an effect on (upon)… 对……有影响;对……起作用

Technical Terms

integrator [ˈintigreitə] *n.* 积分器
damping [ˈdæmpiŋ] *n.* 阻尼
linear ramp 线性斜坡(信号)
log/log coordinates 对数/对数坐标
Bode plot 伯德图
complex－frequency variable 复频率变量
complex plane 复平面
Laplace transform 拉普拉斯变换
angular frequency 角频率
transient response 暂态响应
decaying－exponential response 衰减指数响应
step function 阶跃函数
first－order filter 一阶滤波器
quality factor 品质因子(数)
log [lɔg] *abbr.* logarithm 对数
DC *abbr.* Direct Current 直流电
AC *abbr.* Alternating Current 交流电

Notes

1. 拉普拉斯(Pierre-Simon, marquis de Laplace,1749－1827)是法国天文学家、数学家和物理学家,因研究太阳系的稳定性而闻名于世。

Lesson 6 Analog to Digital Converters

ADCs come in almost as many flavors as ice creams, and at least as much care is needed in choosing the former as is required with the latter.

A popular and readily understood type of ADC is the Flash ADC (Figure 6.1). This is capable of very high-speed conversion and thus can accommodate high sampling rates, but in its basic form is very power hungry [1].

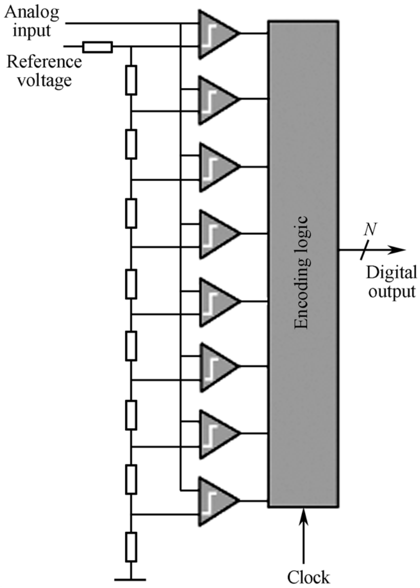


Figure 6.1 Flash ADC Block Diagram

The flash converter operates by simultaneously presenting the input signal to a bank of 2^N-1 comparators, whose reference voltages are set by a resistor chain to exactly correspond to all of the possible sample levels, which can be represented by the converter [2]. The output from each comparator (either a 1 or a 0) is then encoded into an N-bit word representing the input sample level. This approach is the most simple, most intuitive and also the fastest solution for ADC implementation. For large numbers of bits (e.g. >14 bits), the number of resistors needed (2^N-1) becomes prohibitively large for most practical applications. Also, the power consumption is considerably higher than some of the slightly slower and more exotic solutions.

Figure 6.2 shows the output spectrum of a typical flash converter as determined by taking an FFT (Fast Fourier Transform) of the converter output samples for a pure sine wave input. One thing is immediately apparent. The spectrum does not simply consist of the pure input sine wave component, but also has a mass of other components spread throughout the measurement band. These largely arise from the inevitable quantization error (noise), because the converter is trying to represent the analog input level from a finite number of available sample values (dictated by the number of bits in the ADC).

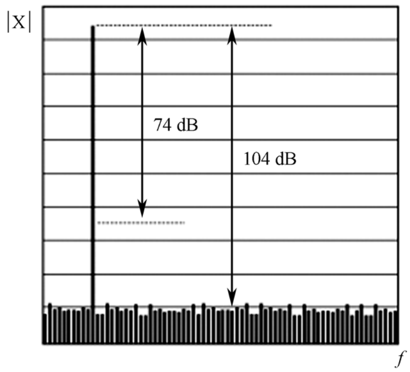


Figure 6.2 FFT of Typical ADC Samples

For our test full-scale sine wave, the effective signal to quantization noise ratio as a function of converter resolution can be readily determined (with a few basic assumptions) to be

$$\mathrm{SNR} = (6.02N + 1.76)\mathrm{dB}$$

where N is the number of bits for the converter.

The converter resolution for the ADC generating the plot in Figure 6.2 is 12 bits, giving a theoretical SNR of 74 dB. Looking at the plot, the difference in levels between the sine wave component and the individual noise components is much nearer 104 dB. The reason for the difference between these two values is that the SNR formula refers to the whole noise contribution, comprising the sum of all the individual noise components making up the FFT[3]. The reason for pointing out this feature is it can be used to increase the effective resolution of a converter by trading off sampling rate as shown below.

Over-sampling to achieve processor gain

Assume that we need to achieve a minimum 70 dB SNR in the conversion process for a given audio application. The formula above suggests a minimum of 12-bit converter

resolution is required (full scale sine wave input and ideal converter), where the noise within the whole band from 0 Hz to $f_s/2$ is included in the measurement. Now, if we employ for example 8×bandwidth sampling (Figure 6.3), we see that the actual audio signal only occupies 1/4 of the base band (0 to $f_s/2$) space whereas the noise is spread uniformly across the band. If we were to now digitally filter this sampled signal, we could remove approximately 3/4 of the noise, increasing the signal to noise ratio by a factor of 4 or 6 dB[4]. This effective increase in SNR is termed the processing gain, achieved by over-sampling the input relative to the 2 ×bandwidth rule.

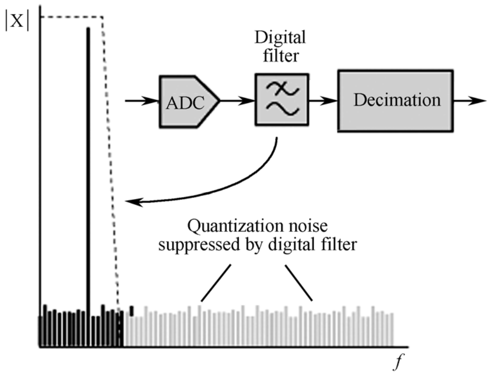


Figure 6.3 Over-Sampling and Quantization Noise

A simple formula for the maximum processing gain can easily be derived: processing gain (dB) = 10 log(sample rate/ 2×signal bandwidth), where it is assumed that digital filtering is employed to restrict the sample bandwidth to exactly match the wanted input signal bandwidth and that the noise is uniformly distributed.

Thus, if we use a 128 times over-sampling design (typically found in minidisk recorders and PC sound cards), we can achieve a real 18 dB improvement in signal to quantization noise, effectively increasing the resolution of the converter from 12 bits to 15 bits. By way of an example, consider using this method to improve the performance of a data converter within a digital cellular phone. The signal bandwidth for a GSM cellular channel is 200 kHz. It is now possible to obtain high speed ADC's with a sampling rate of 80 MSPS and 14-bit resolution, giving an impressive measured 75 dB SNR over the full 0 to $f_s/2$ bandwidth. The processing gain possible is thus: processing gain (dB) = 10 log (80,000,000/2× 200,000) = 26 dB, resulting in a very respectable 75 + 26 = 101 dB SNR for the sampled GSM signal.

Sigma – delta converters

This concept of processing gain leads us nicely into the topic of sigma-delta converters. These take the notion of processing gain to the extreme to achieve very high performance, low cost, low power devices ideally suited for audio applications, with simple analog interfacing (little or no anti-aliasing filters). A block diagram of a basic sigma-delta converter is shown in Figure 6.4.

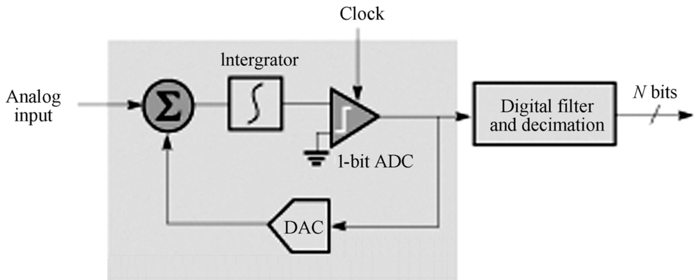


Figure 6.4 Basic Sigma-Delta Converter

The converter is essentially a highly over-sampling 1-bit ADC (the comparator) followed by digital filtering and decimation to realize the processing gain. The effective performance of the converter is greatly enhanced by the addition of circuitry to shape the quantization noise such that, instead of being uniformly spread throughout the 0 to fs/2 band, it is minimized in the band of interest (Figure 6.5).

For a typical 128 × bandwidth over-sampling system, the processing gain alone would give an extra 3 bits of effective resolution (i.e. a 4-bit converter). Noise shaping however can extend this effective resolution much further, with some sigma-delta converters now achieving 24-bit accuracy for audio band applications. The modern converters use a much more sophisticated form of noise shaping processing than that shown in Figure 6.5, which is a simple first-order sigma-delta design, but the basic principle of exploiting processing gain with noise shaping remains the same.

In fact, there are many tens of ADC methods in use, Successive Approximation, Multipass,Interpolating, Subranging, Bit-Per-Stage, to name but a few of the flavors, each potentially having some benefit in performance over its rivals. Luckily, it is not usually necessary to understand how the ADC works in order to make the correct choice of converter for the application. Instead, careful study of the performance specifications on the data sheet will determine the best choice for your application.

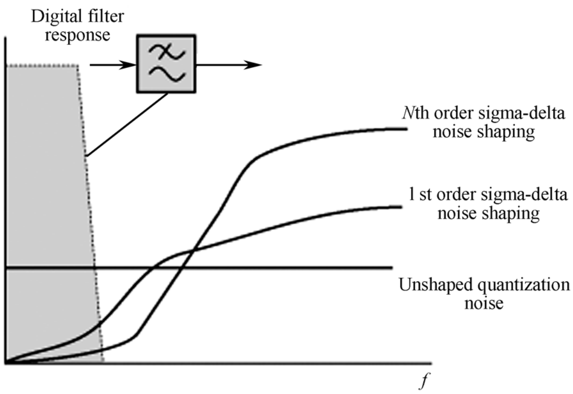


Figure 6.5 Quantization Noise Shaping in Sigma-Delta Converters

New Words

flavor [ˈfleivə] *n.* 风味，滋味
flash [flæʃ] *n.* 闪光，瞬间 *adj.* 迅速的
solution [səˈluːʃn] *n.* 解答，解决办法，解决方案
exotic [igˈzɔtik] *adj.* 特殊的，奇异的，外来的
respectable [risˈpektəbl] *adj.* 相当的
sigma [ˈsigmə] *n.* 西格玛（Σ，σ）
delta [ˈdeltə] *n.* 德尔塔（Δ，δ）
notion [ˈnəuʃn] *n.* 概念，观念，想法
reference [ˈrefrəns] *n.* 参考
approximation [əˌprɔksiˈmeiʃn] *n.* 逼近，近似值

Phrases & Expressions

a mass of 大量的
arise from 由……引起；从……中产生
by way of 经由；作为

Technical Terms

minidisk [ˈminidisk] *n.* 迷你盘

decimation [ˌdesiˈmeiʃn] *n.* 抽取

shaping [ˈʃeipiŋ] *n.* 整形

interpolate [inˈtəːpəleit] *v.* 内插，插值

subrange [ˈsʌbˈreindʒ] *n.* 子区

reference voltage 参考电压

successive approximation 逐次逼近法

GSM *abbr.* Global System for Mobile communications 全球移动通信系统

Notes

1. 此句可译为：闪式 ADC 能够实现高速转换，因而能够提供很高的采样率，但基本形式闪式 ADC 的功耗很大。
2. 此句可译为：闪式 ADC 将输入信号同时置于一列（2^N-1 个）比较器之前，比较器的参考电压是由一组电阻设定的，这些参考电压值和转换器所能代表的采样电压精确对应。
3. 此句可译为：这两个数值存在差异是因为 SNR 公式用的是全部噪声分布，包括了形成 FFT 的全部单个噪声分量之和。
4. 此句含虚拟语气，可译为：假如现在准备对采样信号进行数字滤波，就可以去掉约 3/4 的噪声，从而将信噪比提高 4 dB 或 6 dB。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) What is a filter? A filter is a device that passes electric signals ______ certain frequencies or frequency ranges ______ preventing the passage of others.

Filter circuits are used ______ a wide variety of applications. ______（在电信领域），band-pass filters are used in the audio frequency range (0 kHz to 20 kHz) for ______（调制解调器和语音处理）. High-frequency band-pass filters (several hundred MHz) are used for ______（信道选择）in telephone central offices. Data acquisition systems usually require ______（抗混叠低通滤波器）as well as low-pass noise filters in their preceding ______（信号调理）stages. System power supplies often use band-rejection filters to suppress the 60-Hz line frequency and high frequency transients. In addition, there are filters that do not ______（滤除）any frequencies of a complex input signal, but just add ______（线性相移）to each frequency component, thus contributing a constant ______

(时延). These are called ______(全通滤波器).

At high frequencies (> 1 MHz), all of these filters usually consist ______ passive components ______ inductors (L), resistors (R), and capacitors (C). They are then ______ LRC filters. In the lower frequency range (1 Hz to 1 MHz), however, the inductor value becomes very large and the inductor itself gets quite bulky, ______ economical production difficult. In these ______, ______(有源滤波器) become important. Active filters are circuits that use ______(运算放大器)as the active device ______ combination with some ______(电阻和电容) to provide an LRC-like filter performance at low frequencies.

(2) Choosing the right ______(无源器件)for an analog design is important. In most cases, a right passive component will fit on the same pads ______ a wrong passive component, but not always. Start the design process ______ carefully considering the high frequency characteristics of passive components, and putting the correct part outline on the board ______ the start.

Most designers are totally ______(不懂)the frequency limitations of the passive components they use in ______(模拟电路). Passive components have ______(有限的频率范围), and operation of the part outside of that range can have some very unexpected results. One might think that this discussion only applies ______ high-speed analog circuits. But high frequencies that are ______(辐射或者感应)into a low-speed circuit will affect passive components ______. For example: a simple op amp low-pass filter may well turn into ______(高通滤波器)at RF frequencies.

2. Translate the following passages into Chinese or English.

1) It is extremely hard to define dynamic range (DR) for an op amp, so let's start with a digital-to-analog converter (DAC) where DR is defined as the ratio of the maximum output voltage to the smallest output voltage the DAC can produce.

2) The same definition of DR can be used for an op amp, and the maximum output voltage swing equals V_{OUTMAX}. This output voltage swing is defined as the maximum output voltage the op amp can achieve (V_{OH}) minus the minimum output voltage the op amp can achieve (V_{OL}). V_{OH} and V_{OL} are easily obtainable from an op amp IC data sheet.

3) Noise fluctuates randomly over a period of time, so instantaneous signal or noise levels don't describe the situation adequately. Averages over a long period of time (root mean squared or RMS) are used to describe both the signal and the noise. Signal-to-noise ratio (SNR) was initially established as a measure of the quality of the signal that

exists in the presence of noise. This SNR was a power ratio, and it was established at the output of a circuit. The SNR that we are interested in is a voltage ratio because the impedance is constant, and it is established at the input to the op amp. This means that all noise voltages, including resistor noise voltage, must be calculated in RMS volts at the op amp input.

4) The Chebyshev equal ripple filter distributes the roll-off across the whole passband. It introduces more ripples in the passband but provides a sharper roll-off in the transition region. This type of filter has poorer transient and step responses due to its higher Q values in the stages of the filter.

5) Chip package is the housing that chips come in for plugging into (socket mount) or soldering onto (surface mount) the printed circuit board. Creating a mounting for a chip might seem trivial to the uninitiated, but chip packaging is a huge and complicated industry. The ability to provide more and more I/O interconnections to a die (bare chip) that is increasingly shrinking in size is an ever-present problem. In addition, the smaller size of the package contributes as much to the miniaturization of cell phones and other handheld devices as the shrinking of the semiconductor circuits.

6)电压和水有相似之处——当其供应量很少时,人们才会认识到其价值。低电压系统(此处是指低于 5 V 的单电源供电)让我们认识到电压的价值。

7)在电子技术各领域中,欧姆定律都是基本定律。欧姆定律可表述为 $V = I \cdot R$。欧姆定律可应用于单个器件、一组器件或一个完整电路。当已知流过电路中各部分的电流时,其电压降可由电阻和电流的乘积获得。

8)电路就是无源器件和有源器件的组合。器件按照特定方式排列,以完成期望的功能。器件的排列就形成了一个电路,或叫电路结构。模拟电路设计就是开发各种电路结构。

9)噪声限定了系统可以处理的数据和信号。噪声降低了放大器、接收器等设备辨别信号的能力——因为噪声和输入信号混合起来。运算放大器产生的噪声、电阻噪声和电源噪声决定了能够恢复和测量的信号大小。

10)巴特沃斯(即"最大平坦")滤波器是最常见的通用滤波器。在巴特沃斯滤波器的通带内,衰减单调变化到 3 dB 点(即人们熟知的特征频率点)。不论是几阶巴特沃斯滤波器,这个特征频率点都是相同的。不过,随着滤波器阶数的增加,通带滚降会更接近特征频率点,而在特征频率点和阻带之间的过渡区域滚降也会更陡。

Reading Materials

Passage 1 Filtering? Before or after?

Have you ever needed a lowpass or highpass filter in your circuit and wondered where you should place it in the signal path? Before the introduction of controllers and processors, engineers used analog circuits to implement all filters. With this type of design, you needed to think ahead and work up a "pencil design" before going to the breadboard. If you cut corners, you would probably end up disassembling the circuit and rebuilding it, hoping to get it right the next time. Then came the digital filter. This type of filter implementation can duplicate the frequency response of any analog filter in the digital domain. A major advantage of digital filters is that you can painlessly adjust them with firmware. This scenario sounds too good to be true, and it is.

There are times when you should build the filter with analog hardware and times when it is appropriate to implement it with a controller or processor in firmware. Generally, every circuit in which an analog signal converts to the digital domain should have an analog lowpass filter. This situation is true whether the ADC is a successive-approximation-register (SAR), delta-sigma, pipeline, dualslope, or any other type of converter you may contrive. The placement of this type of filter in the circuit must always be on the analog side, in front of the converter.

Why do you need the analog lowpass filter? Remember that every analog signal has high- and low-frequency noise, whether or not you acknowledge it. The reason you need a lowpass filter goes back to the Nyquist theorem, which illustrates that to accurately convert a signal without contamination, you must first eliminate out-of-band frequencies.

Any signal that passes through the ADC has a magnitude associated with it. The ADC usually faithfully converts the magnitude of that signal as long as the signal frequency is below the converter's input-stage bandwidth. Although the ADC preserves the magnitude, the same is not true for the signal's frequencies. The frequencies above half of the ADC's sampling frequency contaminate the conversion to the point that you won't be able to tell the difference between in-band and out-of-band signals at the converter's output. This phenomenon is known as signal aliasing (Figure 1).

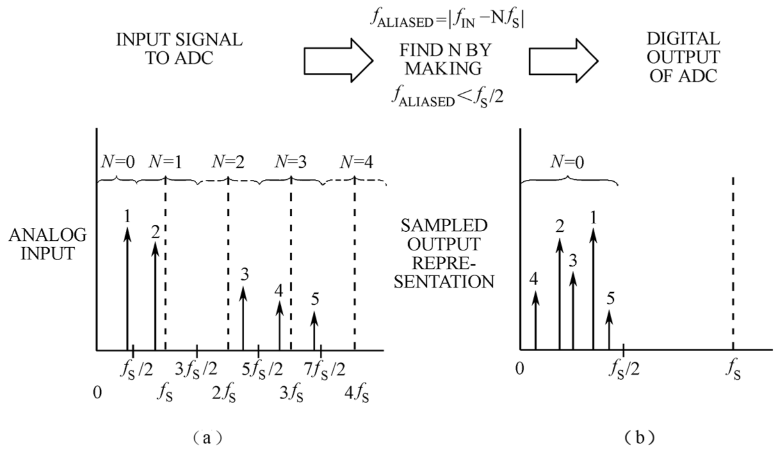


Figure 1 After conversion, the ADC preserves the magnitude of the input signal (a) but aliases the frequencies above half of the sampling frequency, f_S, back below $f_S/2$ (b).

You can see how unwanted noise and signals can permanently embed in the digital signal. Once this contamination occurs, you can't go back and undo it.

Implementing a highpass filter and second lowpass filter is a different matter. You can build these types of filters using analog circuits or firmware. The advantage of building these functions in the controller or processor is that you can implement a variety of easy to adjust filters. This variety includes analog-style filters, such as Butterworth, Bessel, or elliptic. However, you can also implement digital filters, such as an FIR, an IIR, or an FFT. By implementing an FIR filter, you can significantly reduce the in-band noise. You cannot achieve this noise reduction with an analog filter. With an FFT, it is easy to digitally remove unwanted frequencies. You could achieve this removal with an analog circuit, but it would be hardware-intensive.

With digital-filter designs on the horizon, analog filters look fairly unattractive. However, they still have a place in the signal path, as do digital filters. For now, the good news is that digital filters have taken us to the next level of performance, precision, and cost reduction.

Questions:

1) What does the phrase ***cut corners*** mean in the first paragraph?
2) What is firmware? Can you tell its difference from hardware or software?
3) Why does the digital filter "*sounds too good to be true*"?
4) What is ***signal aliasing***?
5) What place does the analog filter have in today's digital world?

Passage 2　Switched-Capacitor Filters

The characteristics of all active filters, regardless of architecture, depend on the accuracy of their RC time constants. Because the typical precision achieved for integrated resistors and capacitors is approximately ±30%, a designer is handicapped when attempting to use absolute values for the components in an integrated filter circuit. The ratio of capacitor values on a chip can be accurately controlled, however, to about one part in 2000. Switched-capacitor filters use these capacitor ratios to achieve precision without the need for precise external components.

In the switched-capacitor integrator shown in Figure 1, the combination of C_1 and the switch simulates a resistor.

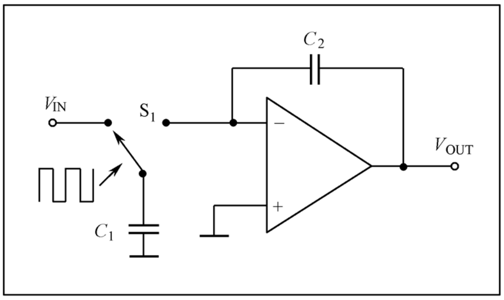


Figure 1　A switched-capacitor integrator

The switch S_1 toggles continuously at a clock frequency f_{CLK}. Capacitor C_1 charges to V_{IN} when S_1 is to the left. When it switches to the right, C_1 dumps charge into the integrator's summing node, from which it flows into the capacitor C_2. The charge on C_1 during each clock cycle is

$$Q = C_1 V_{IN}$$

and thus the average current transferred to the summing junction is

$$I = Qf_C = C_1 V_{IN} \cdot f_C$$

Notice that the current is proportional to V_{IN}, so we have the same effect as a resistor of value

$$R = \frac{V_{IN}}{I} = \frac{1}{C_1 f_C}$$

The integrator's ω_0 is therefore

$$\omega_0 = \frac{1}{RC_2} = \frac{C_1 f_C}{C_2}$$

Because ω_0 is proportional to the ratio of the two capacitors, its value can be controlled with great accuracy. Moreover, the value is proportional to the clock frequency, so you can vary the filter characteristics by changing f_{CLK}, if desired. But the switched capacitor is a sampled-data system and therefore not completely equivalent to the time-continuous RC integrator. The differences, in fact, pose three issues for a designer.

First, the signal passing through a switched capacitor is modulated by the clock frequency. If the input signal contains frequencies near the clock frequency, they can intermodulate and cause spurious output frequencies within the system bandwidth. For many applications, this is not a problem, because the input bandwidth has already been limited to less than half the clock frequency. If not, the switched-capacitor filter must be preceded by an anti-aliasing filter that removes any components of input frequency above half of the clock frequency.

Second, the integrator output (Figure 1) is not a linear ramp, but a series of steps at the clock frequency. There may be small spikes at the step transitions caused by charge injected by the switches. These aberrations may not be a problem if the system bandwidth following the filter is much lower than the clock frequency. Otherwise, you must again add another filter at the output of the switch-capacitor filter to remove the clock ripple.

Third, the behavior of the switched-capacitor filter differs from the ideal, time-continuous model, because the input signal is sampled only once per clock cycle. The filter output deviates from the ideal as the filter's pole frequency approaches the clock frequency, particularly for low values of Q. You can, however, calculate these effects and allow for them during the design process.

Considering the above, it is best to keep the ratio of clock-to-center frequency as large as possible. Typical ratios for switched-capacitor filters range from approximately 28 : 1 to 200 : 1. The MAX262, for example, allows a maximum clock frequency of 4 MHz, so using the minimum ratio of 28 : 1 gives a maximum center frequency of 140 kHz. At the low end, switched-capacitor filters have the advantage that they can handle low frequencies without using uncomfortably large values of R and C. You simply lower the clock frequency.

Questions:

1) What is the meaning of ***handicapped*** in the first paragraph?

2) What benefits can we get when using switched-capacitor filters instead of active filters?

3) What does the word ***toggle*** mean in this article?

4) What function does the switched-capacitor perform?

5) Why is it best to keep the ratio of clock-to-center frequency as large as possible?

Passage 3 Digital to Analog Converters

Figure 1 shows the components that make up the DAC chain. In addition to the conversion of the digital word to a discrete voltage or current level, a zero-order hold is used to "hold" the signal level until the next update. To smooth out the recovered waveform, reconstruction filtering is needed.

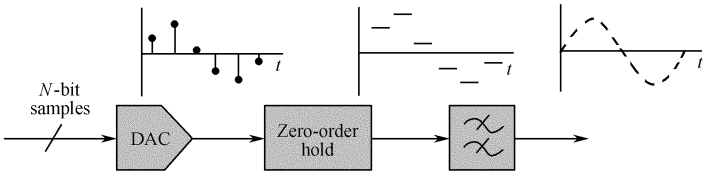


Figure 1 The DAC chain

As we know, over-sampling on the output (providing samples more frequently than is strictly necessary to satisfy the Shannon/Nyquist 2× bandwidth rule) can greatly reduce the specification of the reconstruction filter. We shall consider just two of the many varieties of DAC architectures. The zero-order hold process introduces a small error into the frequency spectrum of the output of the DAC, giving it a sin x/x or sinc weighting. This can be overcome by implementing an "inverse sinc" digital compensation filter prior to the DAC chain. Some DAC devices have an in-built compensation filter so check the data sheet carefully.

The resistor divider DAC

Figure 2 shows the components of the most straightforward DAC, which essentially operates as the reverse of the flash ADC. A known reference voltage is applied across a chain of 2^N-1 resistors such that the voltage measured at any of the tap points exactly corresponds to one of the 2^N available output sample values for an N-bit converter. By

switching in the tap point that corresponds to the value of the digital word to be converted, a basic DAC operation is achieved.

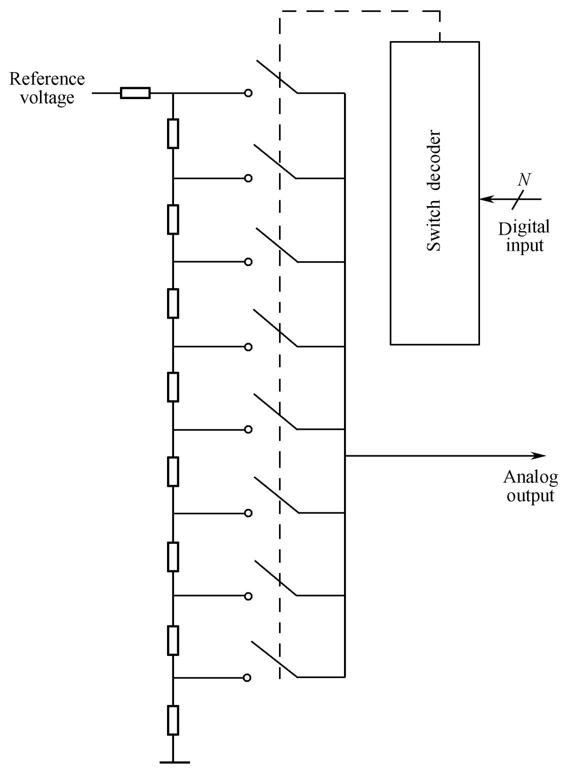


Figure 2 Basic Resistor Divider DAC

The over-sampling approach with a 1-bit DAC is just as effective for output conversion as the 1-bit ADC with over-sampling and noise shaping is for input conversion. A simple sigma-delta DAC is shown in Figure 3.

The first element in the sigma-delta DAC is the interpolation process, which inserts zero samples between each valid sample to realize the increase in data rate, together with sigma-delta DAC digital filtering to perform part of the signal reconstruction. The digital sigma-delta modulator performs a shaping function on the quantization noise, such that in the final reconstructed output, most of the quantization noise is pushed out of the band of interest. The 1-bit DAC is basically a switch, selecting typically a 0 or positive voltage reference level (single supply operation), operating at the very high sampling rate. The analog filter smoothes out the transitions on the output to yield a continuous and high fidelity output waveform.

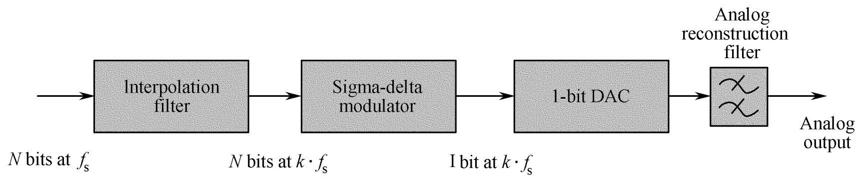


Figure 3　Basic Sigma-Delta DAC

Questions:

1) What components should a DAC chain have?

2) What does Shannon/Nyquist rule say?

3) What's the relation between resistor divider DACs and flash ADCs?

4) What function does the interpolation filter perform in the sigma-delta DAC?

5) How is a basic sigma-delta DAC constructed?

Unit 3

Electronic System Components

Lesson 7 Switching Power Supply

Every new electronic product, except those that are battery powered, requires converting off-line 115 V ac or 230 V ac power to some dc voltage for powering the electronics. Efficient conversion of electrical power is becoming a primary concern to companies and to society as a whole.

Switching power supplies offer not only higher efficiencies but also offer greater flexibility to the designer. Recent advances in semiconductor, magnetic and passive technologies make the switching power supply an ever more popular choice in the power conversion arena today.

Linear versus Switching Power Supplies

Historically, the linear regulator was the primary method of creating a regulated output voltage. It operates by reducing a higher input voltage down to the lower output voltage by linearly controlling the conductivity of a series pass power device in response to changes in its load. This results in a large voltage being placed across the pass unit with the load current flowing through it.

This headroom loss ($V_{drop} \cdot I_{load}$) causes the linear regulator to only be 30 to 50 percent efficient. That means that for each watt delivered to the load, at least a watt has to be dissipated in heat. The cost of the heatsink actually makes the linear regulator uneconomical above 10 watts for small applications. Below that point, however, they are cost effective in step-down applications.

The switching regulator operates the power devices in the full-on and cutoff states. This then results in either large currents being passed through the power devices with a low "on" voltage or no current flowing with high voltage across the device. This results in a much lower power being dissipated within the supply. The average switching power supply exhibits efficiencies of between 70 to 90 percent, regardless of the input voltage.

Higher levels of integration have driven the cost of switching power supplies downward which makes it an attractive choice for output powers greater than 10 watts or where multiple outputs are desired.

Basic Converters

Forward-Mode Converter Fundamentals

The most elementary forward-mode converter is the Buck or Step-down Converter which can be seen in Figure 7.1. Its operation can be seen as having two distinct time periods which occur when the series [1] power switch is on and off. When the power switch is on, the input voltage is connected to the input of the inductor. The output of the inductor is the output voltage, and the rectifier is back-biased. During this period, since there is a constant voltage source connected across the inductor, the inductor current begins to linearly ramp upward which is described by:

$$i_{L(on)} = (V_{in} - V_{out}) \cdot t_{on}/L$$

During the "on" period, energy is being stored within the core material [2] of the inductor in the form of flux. There is sufficient energy stored to carry the requirements of the load during the next off period.

The next period is the "off" period of the power switch. When the power switch turns off, the input voltage of the inductor flies below ground and is clamped at one diode drop below ground by the catch diode. Current now begins to flow through the catch diode thus maintaining the load current loop. This removes the stored energy from the inductor. The inductor current during this time is:

$$i_{L(off)} = (V_{out} - V_D) \cdot t_{off}/L$$

This period ends when the power switch is once again turned on.

Regulation is accomplished by varying the on-to-off duty cycle of the power switch. The relationship which approximately describes its operation is:

$$V_{out} \approx \partial \cdot V_{in}$$

where ∂ is the duty cycle ($\partial = t_{on}/(t_{on} + t_{off})$).

The buck converter is capable of kilowatts of output power, but suffers from one serious shortcoming which would occur if the power switch were to fail short-circuited, the input power source is connected directly to the load circuitry with usually produces catastrophic results. To avoid this situation, a crowbar [3] is placed across the output. A crowbar is a latching SCR which is fired when the output is sensed as entering an overvoltage condition. The buck converter should only be used for board-level regulation.

Flyback or Boost-mode Converter Fundamentals

The most elementary flyback-mode converter is the Boost or Step-up Converter. Its schematic can be seen in Figure 7.2. Its operation can also be broken into two distinct

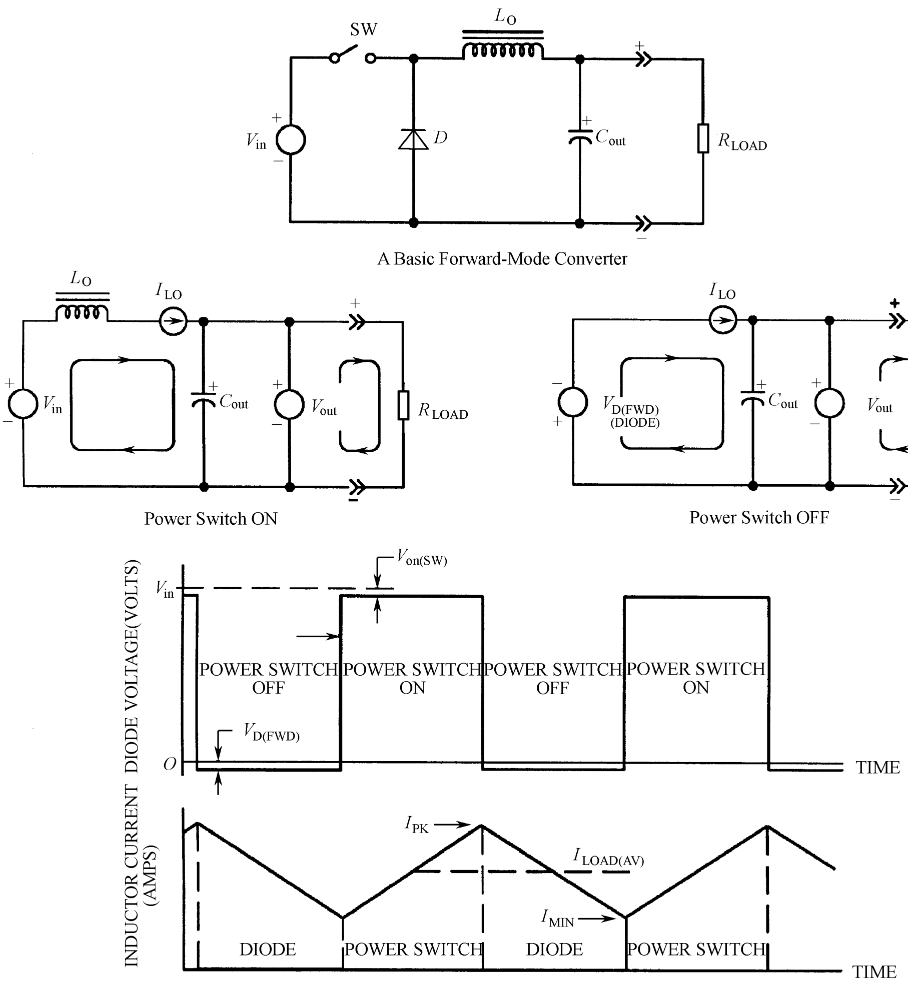


Figure 7. 1 Forward-Mode Converter Operation (Buck Converter Shown)

periods where the power switch is on or off. When the power switch turns on, the input voltage source is placed directly across the inductor. This causes the current to begin linearly ramping upwards from zero and is described by:

$$i_{L(on)} = V_{in} \cdot t_{on}/L$$

Once again, energy is being stored within the core material.

The amount of energy stored during each cycle times the frequency of operation must be higher than the power demands of the load or,

$$P_{sto} = 0.5L \cdot I_{pk}^2 \cdot f_{op} > P_{out}$$

The power switch then turns off and the inductor voltage flys back above the input voltage and is clamped by the rectifier at the output voltage. The current then begins to

linearly ramp downward until the energy within the core is completely depleted. Its waveform which is shown in Figure 7.3 is determined by:

$$i_{L(off)} = (V_{out} - V_{in}) \cdot t_{off}/L$$

The boost converter should also be only used for board-level regulation.

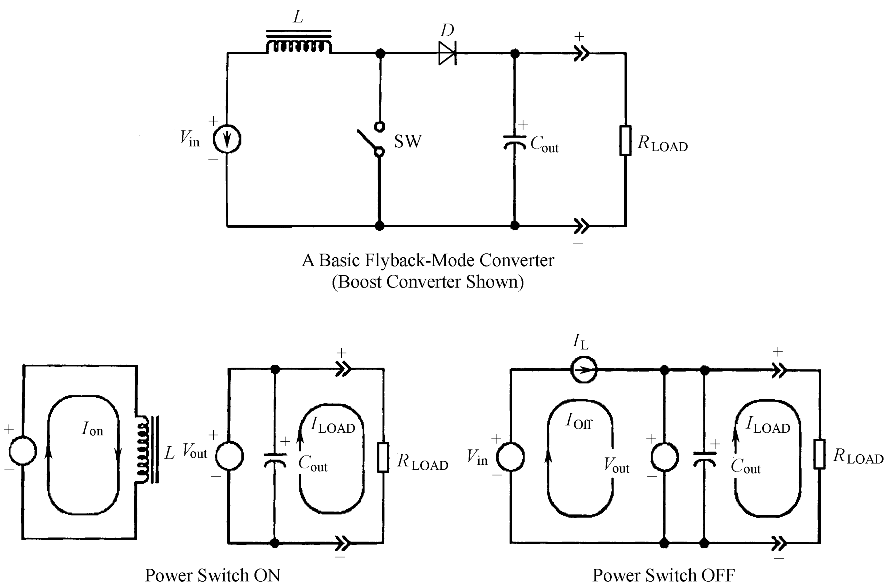


Figure 7.2 Schematic of a Boost Converter

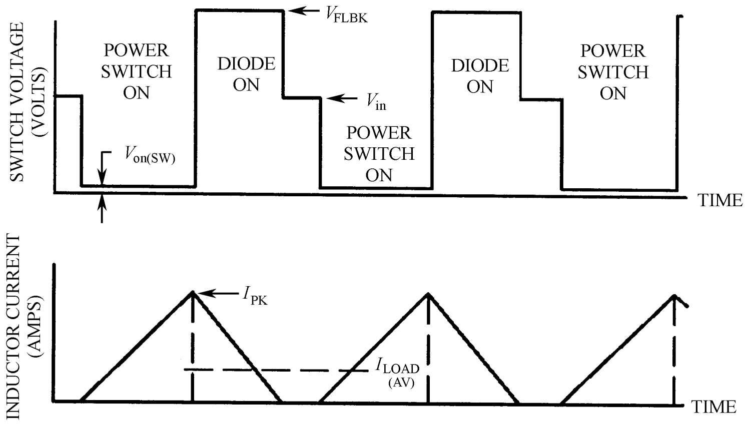


Figure 7.3 Waveforms for a Boost Converter

Topologies

A topology is the arrangement of the power devices and their magnetic elements. Each topology has its own merits within certain applications. Some of the factors which

determine the suitability of a particular topology to a certain application are:

1) Is the topology electrically isolated from the input to the output or not?

2) How much of the input voltage is placed across the inductor or transformer?

3) What is the peak current flowing through the power semiconductors?

4) Are multiple outputs required?

5) How much voltage appears across the power semiconductors?

The first choice that faces the designer is whether to have input to output transformer isolation. Non-isolated switching power supplies are typically used for board-level regulation where a dielectric barrier is provided elsewhere within the system. Non-isolated topologies should also be used where the possibility of a failure does not connect the input power source to the fragile load circuitry. Transformer isolation should be used in all other situations. Associated with that is the need for multiple output voltages. Transformers provide an easy method for adding additional output voltages to the switching power supply. The companies building their own power systems are leaning toward transformer isolation in as many power supplies as possible since it prevents a domino effect during failure conditions.[4]

New Words

arena [ə'riːnə] *n.* 竞技场，舞台

headroom ['hedruːm] *n.* 净空，头上空间，裕量

dissipate ['disipeit] *v.* 耗散

heat-sink ['hiːtsiŋk] *n.* 散热片

uneconomical ['ʌnˌiːkə'nɔmikəl] *adj.* 不经济的，浪费的

cutoff ['kʌtɔːf] *n.* 中止，切断

series ['siəriːz] *adj.* 串联的

biased ['baiəst] *adj.* 加偏压的，有偏向的

flux [flʌks] *n.* 流量，通量

schematic [ski'mætik] *n.* 原理图

topology [tə'pɔlədʒi] *n.* 拓扑，布局

barrier ['bæriə] *n.* 隔板，势垒，阻挡层

fragile ['frædʒail] *adj.* 易被毁坏的

lean [liːn] *vi.* 倾向，偏向

domino ['dɔminəu] *n.* 多米诺骨牌

Phrases & Expressions

domino effect 多米诺效应
failure conditions 故障状态

Technical Terms

switching [ˈswitʃiŋ] *n.* 开关，切换，配电系统
regulator [ˈregjuleitə] *n.* 稳压器
inductor [inˈdʌktə] *n.* 电感器
rectifier [uˈrektifaiə] *n.* 整流器
conductivity [ˌkɔndʌkˈtiviti] *n.* 传导性，传导率
crowbar [ˈkrəubɑː] *n.* 消弧电路，短路器
overvoltage [ˈəuvəˈvəultidʒ] *n.* 超电压，过电压
flyback [ˈflaibæk] *n.* 回扫，回程
boost [buːst] *n.* 升压，放大
dielectric [ˌdaiiˈlektrik] *n.* 电介质，绝缘体
power supply 电源
linear regulator 线性稳压器
load current 负载电流
back bias 反向偏压
SCR *abbr.* Silicon Controlled Rectifier 可控硅整流器

Notes

1. series 一般用作名词，其单数形式和复数形式相同。series 在这里用作形容词。
2. 在本文中，core 是指线圈内或变压器中的铁芯。
3. 短路器(crowbar)是一种保护电子系统免受高电压浪涌破坏的电路。
4. 此句可译为：制造自己系统电源的公司倾向于在尽可能多的电源中采用变压器隔离，因为这种隔离避免了在故障出现时产生的连锁反应。

Lesson 8 Clock Sources

Clock Devices

There are a variety of clock devices available today. Some of them are described below.

Crystals

A crystal is a basic piezoelectric quartz crystal. On its own, it cannot generate electrical clocks. It has to be connected to a clock oscillator to get a clock waveform. There are two kinds of crystals; Series Resonant, which can be modeled as a high Q series L-C circuit, and Parallel Resonant, which can be modeled as a high Q parallel L-C circuit. The series resonant crystal has minimum impedance at the resonating frequency, while the parallel resonant crystal has maximum impedance at the resonating frequency.

Crystal Oscillators

A crystal oscillator is an oscillator with the crystal as the feedback element. There are other kinds of oscillators with active or passive feedback components, but the crystal oscillator provides the most accurate and stable output frequency. Crystal oscillators are the preferred clock source in most high-speed digital systems requiring clocks.

Compensated Oscillators

The output frequency of a crystal oscillator varies with temperature and voltage. Applications that require a highly stable clock usually use compensated oscillators. Compensated Oscillators try to adjust the variation in frequency due to temperature and voltage. Temperature Compensating Oscillators (TCXO) contain circuitry that compensates for temperature changes, and hence combat frequency variations. Oven Controlled Oscillators encase their crystals in a temperature-controlled oven, and so maintain a precise operating temperature at the crystal. Double Oven Oscillators contain two ovens, with the crystal encased in the inner oven, and the temperature control circuitry and the inner oven encased in the outer oven. Such oscillators provide even better temperature stability than Oven Controlled Oscillators. Obviously, as the frequency stability improves, the cost of the oscillator increases.

Voltage Controlled Oscillator

The output of Voltage Controlled Oscillators (VCO) is controlled by a voltage control input pin. Variation between control voltage and frequency is usually nonlinear

over the entire frequency range but is linear within subset ranges.

Frequency Synthesizers

Frequency Synthesizers use one or more Phase-Locked Loops (PLL) to generate one to many different frequencies on their outputs, from one or more reference sources. The reference frequency is usually generated by a crystal attached to the synthesizer. The design goal of frequency synthesizers is to replace multiple oscillators in a system, and hence reduce board space and cost [2]. Figure 8.1 shows a block diagram of a PLL.

A PLL has two inputs, a reference input and a feedback input. A PLL corrects frequency in two ways. The first, frequency correction, corrects large differences in frequency between the reference input and the feedback input. Frequency correction is akin to "rough" tuning and occurs when F_{vco} is less than $0.5F_{ref}$ or greater than $2F_{ref}$. Phase correction is the "fine" tuning and occurs when $0.5F_{vco} < F_{ref} < 2F_{vco}$.

The Phase/Frequency Detector detects differences in phase and frequency between the reference and feedback inputs and generates compensating "Up" and "Down" signals depending on whether the feedback frequency is lagging or leading the reference frequency respectively. These control signals are then passed through a charge pump and a loop filter to generate a control voltage, which controls a VCO. The frequency of this oscillator is dependent on the V_{ctrl} input. At steady state, the VCO frequency is:

$$F_{vco} = F_{ref} \cdot P/Q$$

The output frequency of the PLL can be expressed as

$$F_{out} = (F_{ref} \cdot P)/(Q \cdot N)$$

The *Sample Rate* of a Frequency Synthesizer determines how often the inputs are sampled in order to perform phase and frequency correction. It is expressed as F_{ref}/Q.

The *Acquisition/Lock Time* of a PLL-based Frequency Synthsizer is the amount of time taken by the Frequency Synthesizer to attain the target frequency after power-up, or after a programmed output frequency change.

The *Resolution* of a PLL-based Frequency Synthesizer is based on the number of bits in the P and Q counter. The Resolution will determine in what size increments the frequency can change.

The Deadband of a PLL-based Frequency Synthesizer is the largest phase difference between the reference and the feedback inputs, which will not be corrected by the PLL.

Multiple PLLs are needed within a single frequency synthesizer to generate multiple unrelated frequencies. Frequency synthesizers are gaining in popularity as system complexity increases and systems utilize multiple clocks [3]. The term "Clock Generator" is

interchangeably used with "Frequency Synthesizer."

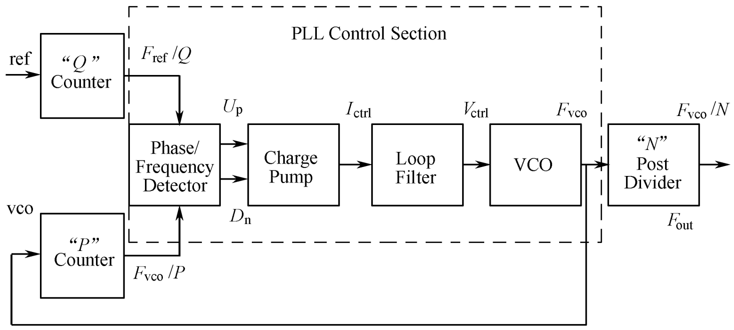


Figure 8.1 Block Diagram of a Phase Locked Loop

Clock Buffers

A Clock Buffer is a device in which the output waveform directly follows the input waveform. The input waveform propagates through the device and is redriven by the output buffers. Hence, such devices have a propagation delay associated with them. In addition, due to the differences between the propagation delay through the device on each input-output path, skew will exist on the outputs.

Clock Parameters

Clock Jitter

Jitter can be defined as the deviations in a clock's output transitions from their ideal positions. The deviation can either be leading or lagging the ideal position. Hence, jitter is expressed in ± ns. Jitter can be classified into three categories: cycle-cycle jitter, period jitter, and long-term jitter.

Cycle-cycle jitter is the difference in a clock's period from one cycle to the next. This kind of jitter is the most difficult to measure and usually requires a Timing Interval Analyzer.

Period jitter, also called short-term jitter, is a change in a clock's output transition from its ideal position over consecutive clock edges. Note that in the case of short-term jitter, the variation of the rising edge of clock from the ideal position is measured and expressed in units of time or frequency.

Long-term jitter is a change in a clock's output transition from its ideal position over "many" cycles. The term "many" depends on the application and the frequency. For PC motherboard and graphics applications, this term "many" usually refers to 10～20 microseconds. For other applications, it may be different.

There are four primary causes of jitter as indicated below:

- Power supply noise
- The internal PLL of the synthesizer
- Random thermal noise from crystal, or any other resonating device
- Random mechanical noise from vibrations of the crystal

Clock jitter affects almost all high-speed synchronous systems. Common applications affected by jitter are PC motherboards, graphics cards, and communications equipment.

Skew

Skew is the variation in arrival time of two signals specified to arrive at the same time. Skew is composed of two parts, the output skew of the driving device, and board design skew, caused by layout variation of board traces.

Clock Driver Skew (*Intrinsic Skew*) is the amount of skew caused by the clock driver itself. There are two kinds of clock driver devices; buffer devices and PLL-based devices. Skew occurs on the output of the buffer devices because of the differences in propagation delay of the input signal through the device. A majority of this difference is attributed to differences in output loading. Skew in PLL-based devices can be very small, since a PLL-based device can be adjusted to compensate for differences in output loading.

Board Design Skew (*Extrinsic skew*) is the amount of skew caused by board layout issues such as:

- Trace Length: The amount of time for a signal to propagate down a trace is dependent on the material of the PCB, length of the trace, width of the trace and capacitive loading. Different trace lengths cause different signal propagation times, and hence cause skew.
- Threshold Voltage Variation: The threshold voltage of the receiving device can cause skew. For example, if a receiving device has a threshold voltage of 1.2 V and another device has a threshold voltage of 1.7 V, and the rise time of the input signal is 1 V/ns, then the two devices will switch 500 ps apart, which is skew.
- Capacitive Loading: The differences in capacitive loading on traces will cause differences in the clock rise times at the load. This affects the time at which the clock edge crosses the input threshold and results in skew.
- Transmission Line Termination: With the extremely fast edge rates in today's clock drivers, traces longer than 4 inches are considered transmission lines. Without proper termination, these lines will exhibit transmission line effects like voltage

reflections, which will cause skew.

Why is skew important? In high-speed systems, clock skew forms an important component of timing margin. A skew of 1 ns is a significant portion of a 15-ns cycle time. If the timing budget does not allow for skew, it is highly likely that the system will perform unreliably.

The simplest method of measuring skew between two outputs of a device is to display both waveforms in a dual-channel oscilloscope and measure the difference between the rising edges. This is the skew.

Tolerance/Accuracy

Tolerance/Accuracy is a measure of how close the part operates to the specified (nominal) frequency, typically referenced at ambient temperature (25℃ +/−5℃). For example, if a part is specified with a 25.000 MHz output, and the long-term (user-defined) average of its output frequency is 25.001 MHz at ambient temperature, the part has +40 ppm (parts per million) accuracy. Frequency tolerance is affected or controlled by controlling the accuracy of the manufacturing and calibrating process for the crystal.

Stability

Stability is a parameter usually associated with crystals and oscillators. Stability is defined as the variation in operating frequency from the ambient temperature frequency (frequency tolerance value) over the operating temperature range and is expressed in ppm. This parameter is specified with a maximum and minimum frequency deviation, expressed in percent or parts per million. Why is stability important? Stability may cause marginal operation of a design over complete temperature range, if it is not accounted for in the design.

Aging

Aging is defined as the systematic change in frequency over time due to internal changes in crystal/oscillator. It is usually expressed in ppm/year, and may be incorporated in the Stability spec, if it is not drawn out separately. It is a parameter usually associated with crystal oscillators. New crystals age faster than old crystals.

Slew

The rate of change of voltage or frequency is called Slew. Slew is usually measured on the rising and falling edges of digital signals. However, rise times and fall times are more commonly specified, instead of slew, in vendor's catalogs. Recently, with the advent of low-power devices, slew is being used to define a rate of change of frequency.

Duty Cycle

Duty Cycle is the ratio of the output high time to the total cycle time. It is expressed as a percentage. 50% is the ideal duty cycle, though most clock manufacturers specify duty cycles from 40%～60%. Duty cycle is important in systems that use both the rising and falling clock edges.

Duty cycles can be expressed for both TTL and CMOS devices. For TTL devices, since the voltage swing is from 0～3 V, the high time is measured at the 1.5 V level. For CMOS devices, since the voltage swing is from 0 - V_{dd} Volts, the high time is measured at $V_{dd}/2$.

New Words

crystal [ˈkristl] *n.* 晶体

piezoelectric [paiˌiːzəuiˈlektrik] *adj.* 压电的

combat [ˈkɔmbæt] *v.* 反对,防止

oven [ˈʌvən] *n.* 恒温箱

encase [inˈkeis] *vt.* 装入,包住

subset [ˈsʌbset] *n.* 子集,附属设备

akin [əˈkin] *adj.* 同族的，类似的

lagging [ˈlægiŋ] *adj.* 滞后的

leading [ˈliːdiŋ] *adj.* 超前的

increment [ˈinkrimənt] *n.* 增加，增量

popularity [ˌpɔpjuˈlæriti] *n.* 普及

propagate [ˈprɔpəgeit] *v.* 传播

deviation [ˌdiːviˈeiʃn] *n.* 偏差,偏移

transition [trænˈziʃn] *n.* 过渡;状态转换

intrinsic [inˈtrinsik] *adj.* 固有的，内在的，本质的

extrinsic [eksˈtrinsik] *adj.* 外在的，外表的，外来的

nominal [ˈnɔminl] *adj.* 名义上的，标称的，额定的

ambient [ˈæmbiənt] *adj.* 周围的 *n.* 周围环境

systematic [ˌsistiˈmætik] *adj.* 系统的,体系的

incorporate [inˈkɔːpəreit] *vt.* 合并

slew [sluː] *n.* 摆动

Phrases & Expressions

on one's own 独自地，主动地
gain in 在……方面增加
allow for 虑及，体谅
draw out 取出，抽出

Technical Terms

oscillator [ˈɔsileitə] *n.* 振荡器
resonant [ˈrezənənt] *adj.* 谐振的
tuning [ˈtjuːniŋ] *n.* 调谐
detector [diˈtektə] *n.* 检测器，检波器
skew [skjuː] *n.* 相位偏移，时滞
termination [ˌtəːmiˈneiʃn] *n.* 端接法
series resonant 串联谐振
parallel resonant 并联谐振
resonating frequency 谐振频率
feedback element 反馈元件
frequency synthesizer 频率合成器
phase detector 鉴相器
frequency detector 鉴频器
charge pump 电荷泵
loop filter 环路滤波器
deadband 死区
propagation delay 传播延迟
clock jitter 时钟抖动
transmission line 传输线
voltage reflection 电压反射
duty cycle 占空比
TCXO *abbr.* Temperature Compensated Crystal Oscillator 温度补偿晶体振荡器
VCO *abbr.* Voltage Controlled Oscillator 压控振荡器
PLL *abbr.* Phase Locked Loop 锁相环

Notes

1. 此句可译为：设计频率合成器的目的是用以替代系统中的多个振荡器，从而减少了电路板空间、降低了系统成本。
2. 此句可译为：随着系统复杂性的提高和多个时钟在系统中的使用，频率合成器应用得越来越普遍。词组“gain in”主要有两个意思：“增加、增长”（to increase; grow）和“改善 、提高”（to become better; improve）。

Lesson 9 Interconnect

Interconnect devices are used for interconnecting components, circuit boards, and modules. The needs for these devices must be considered during the design phase.

Interconnect devices must meet certain performance requirements in most or all of the following categories:

- Signal integrity
- Power loss
- Electrical characteristics: contact resistance, inductance and capacitance; voltage and power ratings; shielding; filtering
- Mechanical characteristics: contact count, contact spacing, contact forces; shock and vibration; size; soldering technique
- Environmental issues: cleaning materials; thermal; corrosion protection

Signal integrity is defined by loss of quality of the input signal. Ideally, the signal at the output of an interconnect should be equal in all characteristics with the signal at the input of the interconnect. In reality, signal degradation occurs. The user must define what level of degradation is acceptable or, alternatively, must define minimum acceptable signal levels at the output. Tests that indicate signal integrity include:

- Voltage standing wave ratio (VSWR)[1]
- Frequency response
- Rise-time of signal edges
- Current flow

High-Speed Connector Problems

In today's designs, with clock rates over 100 MHz and rise times commonly 1 nanosecond (ns) or less, designers cannot ignore the role interconnections play in a logic design. Interconnect effects can play a significant part in the timing and noise characteristics of a circuit. The faster clock rates and rise times increase both capacitive and inductive coupling effects, which makes crosstalk[2] problems greater. They also mean shorter time for reflections to decay before the data is clocked and read, which decreases the maximum line length that can be used for unterminated systems. This all means that one of the major interconnect challenges is to ensure signal integrity as the high-speed pulses move along the total interconnect path, from device to PCB, through the PCB to the backplane, and on out to any network connections which may be present.

An interconnect that must pass the short rise time of a high-speed signal pulse can be detrimental to maintaining signal integrity due to an unwanted reflection. The shorter the rise time, the greater the risk of degradation of the signal. The ideal situation is that the connector will have the appropriate termination characteristics to create no degradation of the signal.

Crosstalk

In addition to general signal degradation, another potential impact of a connector on signal integrity is increased crosstalk, which is caused by inductive and capacitive coupling between contacts and/or lines feeding contacts. This can cause noise to be induced in one contact/line from a switching signal in another contact/line. Crosstalk can be expressed in a percentage of the voltage swing of the signal, or in decibels (dB), where 0 dB represents 100% crosstalk, −20 dB represents 10% crosstalk, and −40 dB represents 1% crosstalk.

Two types of crosstalk can be generated: backward crosstalk and forward crosstalk. Backward, also called near-end crosstalk, is measured at the driving end of the connector and represents the sum of the capacitive and inductive coupling. Forward, also called far-end crosstalk, is measured at the receiving end of the connector and represents capacitive minus inductive coupling.

Crosstalk can also be single-line or multiline. Single-line crosstalk is the result of one driven line inducing a voltage on one quiet line. Multiline crosstalk is two or more driven lines inducing voltage(s) on one quiet line. In most systems, many lines/contacts

will be in use at once, so multiline crosstalk is a more realistic figure.

To completely understand the crosstalk specifications of a connector, a user needs the following information:

1. Is the crosstalk specification measured on the actual connector, or is it the result of a simulation?

2. What is the value of the signal rise time used for the crosstalk measurement, and the digital technology used to generate it (TTL and CMOS rise times are typically 10% to 90% rise times, while ECL is typically a 20% to 80% rise time)?

3. What is the signal pattern on the connector, i.e., the location of the ground lines relative to the measured lines?

4. How many lines are being driven by the switched signal?

5. What are the source and load impedances?

6. Is the crosstalk specification for forward or backward, or the total?

7. Is the crosstalk specification for single-line or multiline?

8. If multiline, how many lines are being driven?

One common way to minimize crosstalk is to intersperse signal contacts with ground contacts. However, this has the disadvantage of minimizing signal contact density.

Levels of Interconnects

Granitz has defined six levels of interconnects:

Level 1: chip pad to package leads, e.g., wire bonds

Level 2: components to circuit board, e.g., PLCC socket

Level 3: circuit board (board-to-board) connector, e.g., edge-card for motherboard to daughter board[3] connection

Level 4: subassembly interconnects, e.g., ribbon cables

Level 5: subassembly to I/O, e.g., BNC connector[4]

Level 6: system interconnects, e.g., Ethernet[5] connectors

Each level of interconnection may be accomplished in several ways. For example, a level 4 interconnect between circuit boards may use a card edge connector, a ribbon cable, or even discrete wires.

New Words

shielding [ˈʃiːldiŋ] *adj.* 防护的,屏蔽的

spacing [ˈspeisiŋ] *n.* 间隔,间距

soldering [ˈsɔldəriŋ] *n.* 焊接
degradation [ˌdegrəˈdeiʃn] *n.* 降级，退化
backplane [ˈbækplein] *n.* 背板
detrimental [ˌdetriˈmentl] *adj.* 有害的
intersperse [ˌintəˈspəːs] *vt.* 散布，点缀
subassembly [ˈsʌbəˈsembli] *n.* 部件，组件，配件

Phrases & Expressions

in reality 实际上，事实上
in addition to 除……外

Technical Terms

capacitive [kəˈpæsitiv] *adj.* 电容性的
inductive [inˈdʌktiv] *adj.* 电感性的
reflection [riˈflekʃn] *n.* 反射
decibel [ˈdesibel] *n.* 分贝
feeding [ˈfiːdiŋ] *n.* 馈电，输送
crosstalk [ˈkrɔsˌtɔːk] *n.* 串扰
pad [pæd] *n.* 焊点，焊盘
lead [liːd] *n.* 引线
bond [bɔnd] *n.* 接头
ethernet [ˈiːθənet] *n.* 以太网
power loss 功率损耗
contact resistance 接触电阻
corrosion protection 防腐蚀
frequency response 频率响应
current flow 电流
edge card 边缘卡
motherboard 母板
daughter board 子板
ribbon cable 扁平电缆
VSWR *abbr.* Voltage Standing Wave Ratio 电压驻波比
BNC *abbr.* Bayonet Neill－Concelman 同轴电缆卡环形接头

Notes

1. “电压驻波比”(VSWR)是用来度量传输线和负载之间阻抗失配的；电压驻波比越大，阻抗失配越严重。电压驻波比的理想值(最小值)为1，即阻抗完全匹配。
2. “串扰”(crosstalk)是指由于泄漏或耦合造成的、出现在通信信道上的其他电路传送过来的信号。
3. 子板(daughter board)是指插入母板(mother board)中的印制电路板。
4. BNC连接器即“同轴电缆卡环形接头”，由Bayonet Neill－Concelman发明。
5. 以太网(Ethernet)是当前应用最普遍的局域网组网技术，其技术标准由IEEE 802.3给出。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) Printed circuit board is a flat board that holds chips and other electronic components. The board is made of ______(层)(typically 2 to 10) that interconnects components ______(经由)copper pathways. The main printed circuit board in a system is called a “system board” or “motherboard,” ______ smaller ones that plug into the slots in the main board are called “boards” or “cards.”

The printed circuit board of the 1960s connected ______(分立器件)together. The circuit board of the 1990s interconnects chips, each containing ______(几十万个)or millions of elementary components. The “printed” circuit is really an ______(蚀刻电路). A ______(铜箔)is placed over the ______(玻璃纤维)or plastic base of each layer and covered with a photoresist. Light is shined through a ______(负像)of the circuit paths onto the photoresist, hardening the areas that will remain ______ etching. When passed through an acid bath, the unhardened areas are washed away. The finished layers are then glued together. A ______(相似的处理)creates the microminiaturized circuits ______ a chip.

(2) The interface is a connection and interaction between hardware, software and the user. Hardware interfaces are the plugs, sockets, wires and the ______(电脉冲) traveling through them ______ a particular pattern. Also included are electrical timing considerations. Examples ______ RS-232 transmission, the Ethernet and Token Ring network topologies and the IDE, ESDI, SCSI, ISA, EISA and Micro Channel

interfaces. Software interfaces are the languages, codes and messages programs use ______ communicate with each other and to the hardware. Examples are the applications that run under the Mac, DOS and Windows operating systems ______ the SMTP e-mail and LU 6.2 ______(通信协议). User interfaces are the keyboards, mice, commands and menus used for communication between you and the computer. Examples are the command lines in DOS and UNIX, and the Mac, Windows and Motif ______(图形界面).

Interfacing is a major part of what ______(工程技术人员、程序设计人员和咨询人员)do. Users "talk to" the software. The software "talks to" the hardware and other software. Hardware "talks to" other hardware. All this is interfacing. Every interface ______(意味着一种结构). Electrical signals are made up of ______(电压电平、频率和持续时间). The data passed from one device or program to another has a precise format (header, body, trailer, etc.). Every interface implies a function. At the ______ level, electronic signals activate functions; data are ______, written, ______, received, analyzed for error, etc. At the software level, instructions ______ the hardware (access methods, data link protocols, etc.). At higher levels, the data transferred or transmitted may itself request functions to be performed (client/server, program to program, etc.).

2. Translate the following passages into Chinese or English.

1) Use of custom gate array logic, application specific integrated circuits (ASICs), ball grid arrays (BGAs), multichip modules (MCMs), and digital devices operating in the subnanosecond range present new and challenging opportunities for EMC engineers.

2) A resistor at high frequency acts like a series combination of inductance with the resistor in parallel with a capacitor. A capacitor at high frequency acts like an inductor and resistor in series combination with the capacitor plates.

3) SPICE (Simulation Program with Integrated Circuit Emphasis) is a program widely used to simulate the performance of analog electronic systems and mixed mode analog and digital systems. SPICE solves sets of non-linear differential equations in the frequency domain, steady state and time domain and can simulate the behavior of transistor and gate designs.

4) IBIS (I/O Buffer Information Specification) is a format for defining the analog characteristics of the input and output of integrated circuits. IBIS models are ASCII files that provide the behavioral information required to model the device without divulging the proprietary design of the circuit.

5）In multiconductor systems，excessive line-to-line coupling，or crosstalk，can cause two detrimental effects. First，crosstalk will change the performance of the transmission lines in a bus by modifying the effective characteristic impedance and propagation velocity. Additionally，crosstalk will induce noise onto other lines. These aspects of crosstalk make system performance heavily dependent on data patterns，line-to-line spacing，and switching rates.

6）晶体是一种含有统一排列分子的透明石英材料。

7）随着高速数字系统变得越来越复杂，时序和信号完整性的实现也越来越困难。

8）串扰是两条导线间的能量耦合。当不同结构的电磁场相互作用时，串扰就会发生。

9）在数字系统中，串扰是非常普遍的。在芯片上、印制电路板上、芯片封装上和连接电缆上都可能出现串扰。

10）许多系统工作在高频上，导体在这些频率上不再仅仅表现为导线，而是展现出高频效应并表现为传输线。如果没有正确处理这些传输线，就可能在无意中破坏了系统时序。

Reading Materials

Passage 1 Some Circuit Board Layout Techniques

Noise Sources

Noise is the primary limitation on analog circuitry performance. Some types of noise include:

• Conducted Emissions — noise that the analog circuitry generates through its connections to other circuits. This is usually negligible in analog circuitry, unless it is high power (such as an audio amplifier that draws heavy currents from its power supply).

• Radiated Emissions — noise that the analog circuitry generates, or transmits, through the air. This is also usually negligible in analog circuitry, unless it is high frequency such as video.

• Conducted Susceptibility — noise from external circuitry that is conducted into the analog circuit through its connections to other circuits. Analog circuitry must be connected to the "outside world" by at least a ground connection, a power connection, an input, and an output. Noise can be conducted into the circuit through all of these paths, as well as any others that are present.

• Radiated Susceptibility — noise that is received through the air (or transmitted into the analog circuitry) from external sources. Analog circuitry, in many cases, resides on a PCB that may have high-speed digital logic including DSP chips. High-speed clocks and switching digital signals create considerable radio frequency interference (RFI). Other sources of radiated noise are endless: the switching power supply in a digital system, cellular telephones, broadcast radio and TV, fluorescent lighting, nearby PCs, lightning in thunderstorms, and so on. Even if the analog circuitry is primarily audio in frequency, RFI may produce noticeable noise in the output.

How Many Layers are Best?

Depending on the complexity of the overall circuitry being designed, a designer must decide how many layers the PCB should be.

Single-Sided

Very simple consumer electronics are sometimes fabricated on single-sided PCBs,

keeping the raw board material inexpensive with thin copper cladding. These designs frequently include many jumper wires, simulating the circuit routing on a doublesided board. This technique is only recommended for low-frequency circuitry. For reasons described below, this type of design is extremely susceptible to radiated noise. It is harder to design a board of this type, because of the many things can go wrong. Many complex designs have been successfully implemented with this technique, but they require a lot of forethought. An example is a television set that puts all of the analog circuitry on a single-sided board at the bottom of the case, and uses the metallized CRT itself to shield the board from a separate digital tuning board near the top of the set. Be prepared to get creative if the design demands high volume, low cost PCBs. If a single-sided PCB is a requirement, remember the trace resistance!

Double-Sided

The next level of complexity is double-sided. Doubled-sided boards are easier to route because there are two layers of foil, and it is possible to route signals by crossing traces on different layers. Crossing traces, however, is not recommended for analog circuitry. Wherever possible, the bottom layer should be devoted to a ground plane, and all other signals routed on the top layer. A ground plane provides several benefits:

- Ground is frequently the most common connection in the circuit. Having it continuous on the bottom layer usually makes the most sense for circuit routing.
- It increases mechanical strength of the board.
- It lowers the impedance of all ground connections in the circuit, which reduces undesirable conducted noise.
- It adds a distributed capacitance to every net in the circuit - helping to suppress radiated noise.
- It acts a shield to radiated noise coming from underneath the board.

Multi-Layer

Double-sided boards, in spite of their benefits, are not the best method of construction, especially for sensitive or high-speed designs. The most common board thickness is 1.5 mm. This separation is too great for full realization of some of the benefits listed above. Distributed capacitance, for example, is very low due to the separation. Critical designs call for multi-layer boards. Some of the reasons are obvious:

- Better routing for power as well as ground connections. If the power is also on a plane, it is available to all points in the circuit simply by adding vias.
- Other layers are available for signal routing, making routing easier.
- There will be distributed capacitance between the power and ground planes,

reducing high frequency noise.

There are other reasons for multi-layer boards, however, that may not be obvious or intuitive.

• Better EMI/RFI rejection. There is due to the image plane effect, which has been known since the time of Marconi. When a conductor is placed close to a parallel conductive surface, most of the high frequency currents will return directly under the conductor, flowing in the opposite direction. This mirror image of the conductor within the plane creates a transmission line. Since currents are equal and opposite in the transmission line, it is relatively immune to radiated noise. It also couples the signal very efficiently. The image plane effect works equally well with ground and power planes, but they must be continuous. Any gap or discontinuity causes the beneficial effects to quickly vanish.

• Reduced overall project cost for small production runs. Although multi-layer boards are more expensive to manufacture, EMI/RFI requirements from the FCC or other agencies may require expensive testing of the design. If there are problems, it can force a complete redesign of the PCB, leading to additional rounds of testing. A multi-layer PCB can have as much as 20-dB better EMI/RFI performance over a 2-layer PCB.

Grounding

Good grounding is a system-level design consideration. It should be planned into the product from the first conceptual design reviews.

The Most Important Rule: Keep Grounds Separate.

Separate grounding for analog and digital portions of circuitry is one of the simplest and most effective methods of noise suppression. One or more layers on multi-layer PCBs are usually devoted to ground planes. If the designer is not careful, the analog circuitry will be connected directly to these ground planes. The analog circuitry return, after all, is the same net in the netlist as digital return. Autorouters respond accordingly and connect all of the grounds together, creating a disaster. After the fact separation of grounds on a mixed digital and analog board is almost impossible. Every ground connection in the analog circuitry must be lifted from the board and connected together. For surface mount boards, this results in a colossal mess of "tombstoned" passive components and floating IC leads.

Other Ground Rules

• Ground and power planes are at the same ac potential, due to decoupling capacitors and distributed capacitance. Therefore, it is important to isolate the power

planes as well.

• Do not overlap digital and analog planes. Place analog power coincident with analog ground, and digital power coincident with digital ground. If any portion of analog and digital planes overlap, the distributed capacitance between the overlapping portions will couple high-speed digital noise into the analog circuitry. This defeats the purpose of isolated planes.

Separate grounds does not mean that the grounds are electrically separate in the system. They have to be common at some point, preferably a single, low-impedance point. System-wise, there is only one ground, and it is the electrical safety ground in an ac-powered system or battery ground in a dc-powered system. Everything else "returns" to that ground. It would be a good idea to develop the discipline to refer to everything that is not a ground as a return. All returns should be connected together at a single point, which is system ground. At some point, this will be the chassis. It is important to avoid ground loops by multiple connections to the chassis. Ensuring only one chassis ground point is one of the most difficult aspects of system design.

• If at all possible, dedicate separate connector pins to separate returns, and combine the returns only at system ground. Aging and repeated mating causes connector pins to increase in contact resistance, so several pins are needed. Many digital boards consist of many layers and hundreds or thousands of nets. The addition of one more net is seldom an issue, but the addition of several connector pins almost always is. If this cannot be done, then it will be necessary to make the two returns a single net on the PCB - with very special routing precautions.

• It is important to keep digital signals away from analog portions of the circuit. It makes little sense to isolate planes, keep analog traces short, and place passive components carefully if there are high-speed digital traces running right next to the sensitive analog traces. Digital signals must be routed around analog circuitry, and not overlap analog ground and power planes. If not, the design will include a new schematic symbol - the broadcasting antenna! Most digital clocks are high enough in frequency that even small capacitances between traces and planes can couple significant noise. Remember that it is not only the fundamental frequency of the clock that can cause a potential problem, but also the higher frequency harmonics.

• It is a good idea to locate analog circuitry as close as possible to the I/O connections of the board. Digital designers, used to high current ICs, will be tempted to make a 50-mil trace run several inches to the analog circuitry thinking that reducing the resistance in the trace will help get rid of noise. What they have actually done is create a

long, skinny capacitor that couples noise from digital ground and power planes into the op amp, making the problem worse!

Questions:

1) Can you tell some types of noise that usually cannot be negligible in analog circuitry?

2) Why is single-sided PCB only recommended for low-frequency circuitry?

3) What benefits does a ground plane provide?

4) Can you tell some reasons why critical designs call for multi-layer boards?

5) Do you know any important grounding rules?

Passage 2 Choosing the Right Power-Supply IC

The chief purpose of most power-supply ICs is to regulate. These devices take an unregulated input voltage and provide a regulated output voltage, that is, an output voltage that remains steady despite varying input voltage or output current. This accounts for the names linear regulator and switching regulator. The exception is the charge pump: Depending on the specific device, a charge pump's output can be either regulated or unregulated.

Linear Regulators

Linear regulators are often the smallest, usually the least expensive, and always the least noisy of the various types of power-supply ICs. See Figure 1. A linear regulator both steps down and regulates the voltage supplied to it with a minimal number of external components. Because these devices contain no switching elements, they generate little noise. Also, the circuit-board layout of linear regulators is less critical than for switching regulators and charge pumps.

Why would you use any power-supply IC other than a linear regulator? One reason: A linear regulator can only provide an output voltage that's smaller than its input voltage. If you wanted to create a voltage that's higher than the input voltage or of opposite polarity, you'd have no choice but to use a switching regulator or a charge pump.

Another reason: Efficiency. Converting a voltage to another voltage always wastes power. In the ideal case where a regulator wastes no power, its efficiency rating would

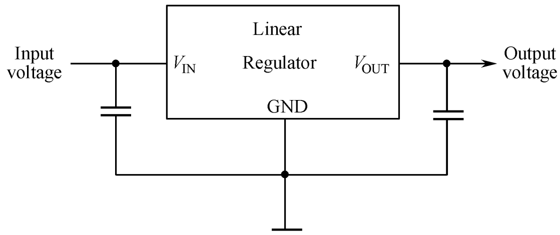


Figure 1　A linear regulator

be 100%. If half the power supplied to a regulator found its way to the regulator's load, its efficiency would be 50%.

A linear regulator is usually, though not always, less efficient than a switching regulator. You can calculate a linear regulator's efficiency by dividing its output voltage by its input voltage. (This formula is sufficiently accurate if the current that powers the regulator-the regulator's supply or quiescent current-is a small percentage of the current drawn from the regulator's output, and in most cases it is.) Thus, when the voltage of the source powering a linear regulator is near the regulator's output voltage, efficiency is high; in that case, a linear regulator may be a better choice than a switching regulator.

A high-efficiency regulator provides a distinct advantage in portable equipment, as less wasted power results in longer battery life. You may need a high-efficiency regulator for another reason: Wasted power is dissipated as heat. Thus, a high-efficiency power supply often suits wall-powered equipment as well as portable equipment. It can reduce the temperature within an enclosure to a tolerable level in either case.

Switching Regulators

Switching regulators share none of the advantages of linear regulators. Switching regulators consume more board area (except perhaps when a linear regulator requires a heatsink to dissipate the power lost within it), cost more, and generate more noise than their linear counterparts. Yet for years switching regulators have been enormously popular with power-supply designers. Why? The reason is because these devices boast excellent efficiency when subjected to many combinations of input voltage and load current (as high as 96% for both step-up and step-down switchers, although a step-down is typically more efficient, and up to 90% for an inverter). Also, if you need to step up, step down, or invert a voltage, you'll find that switching regulators are the

only devices capable of these operations for load currents above approximately 125 mA. You can use charge pumps to perform these operations, but the load currents these devices allow are limited. It's simply too expensive to make the switches internal to charge pumps large enough to handle load currents above the 125mA level just mentioned, although a few charge pumps supply several hundred milliamps.

Switching regulators are so named because they switch a power transistor, which, when used in conjunction with an inductor, efficiently converts one voltage to another. See Figure 2. When these power transistors switch, they do so very quickly, as these fast transitions improve the regulator's efficiency. To understand why, first consider the power transistor's power dissipation when it isn't transitioning. When the transistor is off, voltage appears across it, but no current flows through it; thus, no power is lost. When the transistor is on, a small voltage appears across it, while appreciable current may flow through it; thus, typically, a small amount of power is lost. When the power transistor transitions from an off state to an on state, or vice versa, voltage appears across the transistor, while current flows through it; therefore, appreciable power may be lost. Speeding up the switching process reduces these transition losses.

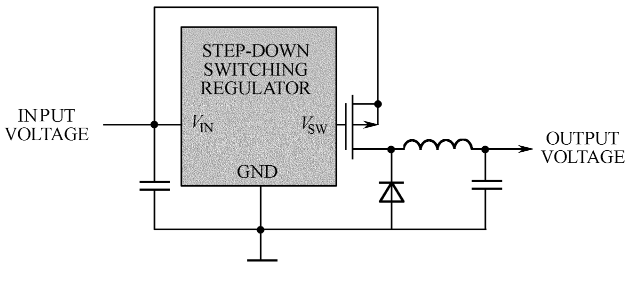


Figure 2a Step-down switching regulators

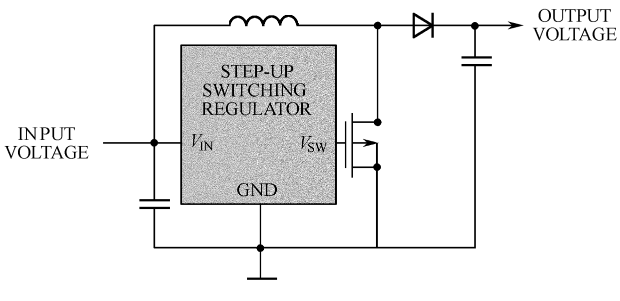


Figure 2b Step-up switching regulators

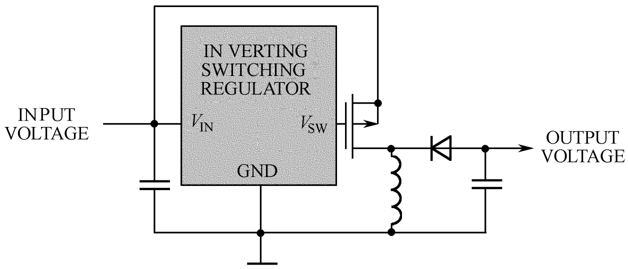


Figure 2c　Inverting switching regulators

These fast transitions, along with the heavy currents that often flow in these circuits, make circuit-board layout critical. Because switching-regulator circuits require a well-thought-out layout and because the components external to the switching-regulator IC must be specified correctly, of the various types of power supplies switching regulators require the most careful design.

Charge Pumps

Charge pumps constitute the least-known category of the three types of power-supply ICs discussed here. These devices perform the same functions as switching regulators, but without an inductor. Instead, charge pumps use capacitors to step down, invert, or boost the voltages that power them. See Figure 3.

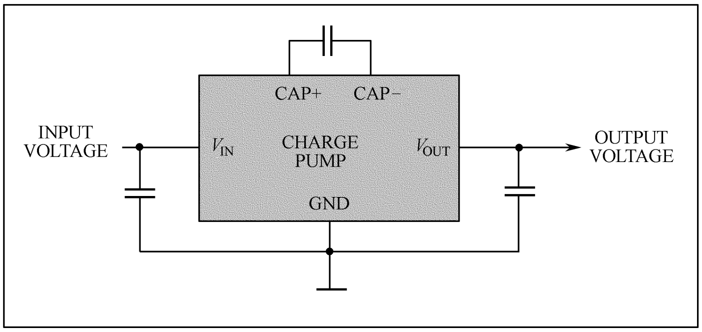


Figure 3　A charge pump

Charge pumps come with both unregulated and regulated outputs. With an unregulated charge pump, as a circuit connected to its output draws more and more current, its output voltage drops proportionately. The charge pump's output impedance is thus essentially a fixed resistance. Unregulated charge pumps, when used in the inverting mode, provide an output voltage equal to the voltage powering the device, but

with opposite polarity. As load current increases, the magnitude of this voltage drops, as discussed above. When used in the doubling mode, these charge pumps precisely double the applied voltage, and the output voltage also drops as load current increases.

Regulated charge pumps can step up, step down, or invert the applied voltage. Unlike unregulated charge pumps, these devices provide output-voltage levels that aren't strictly dependent on the voltage level fed to them. Thus, these devices could, for example, create a 5 V output from a 3.3 V input. Also, because they're regulated, as the output current increases, the output voltage remains essentially constant. As mentioned above, the amount of current that can be drawn from these devices, as well as from unregulated charge pumps, is limited; the upper end is about 125 mA, although there are a few parts that handle several hundred milliamps. It's not economical to build charge pumps that supply large load currents. Instead, inductor-based switching regulators are well suited for this situation.

A charge pump switches the capacitors connected to it, and thus creates noise. For several reasons, this noise is usually of smaller magnitude than a switching regulator's noise. For one, load currents are lighter. Also, because these circuits don't include an inductor, no magnetic noise is created. Finally, when a charge pump interrupts the current flowing through a capacitor connected to it, a voltage spike isn't created. A switching regulator interrupting the current flowing through an inductor usually creates a voltage spike.

Questions:

1) What is the meaning of ***regulator*** in this article?

2) Why linear regulators generate little noise?

3) Why are linear regulators less popular than switching regulators or charge pumps?

4) Which type of power supply requires the most careful design? Why?

5) What kind of components does a charge pump use to perform voltage regulation?

Passage 3 Specifying Quartz Crystals

Quartz crystals are available in a myriad of shapes and sizes, and can range widely in performance specifications. These specifications include resonance frequency, resonance mode, load capacitance, series resistance, holder capacitance, motional

inductance and capacitance, and drive level. Understanding these parameters and how they relate to the crystal's performance will allow you to successfully specify crystals for your circuit application.

A quartz crystal can be modeled as a series LRC circuit in parallel with a shunt capacitor. Figure 1 shows this generic circuit model.

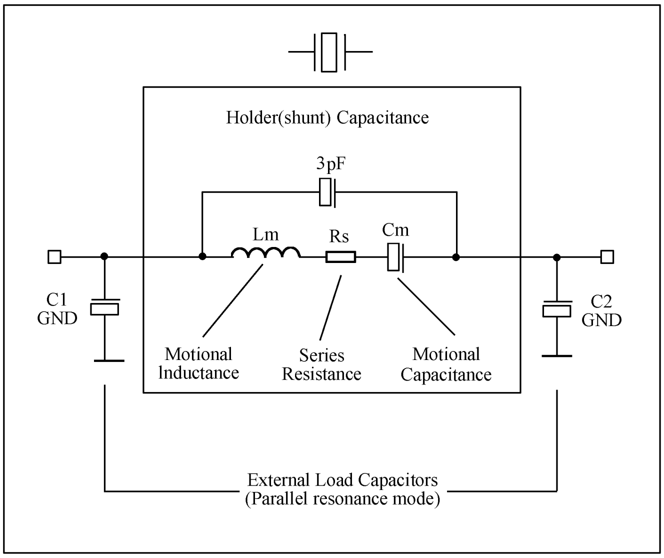


Figure 1 Generic crystal model (fundamental mode)

Now let's look at each key performance specification in detail.

Resonance Frequency

Crystals below 30 MHz are often specified at the fundamental frequency, but above 30 MHz they are typically specified as 3rd, 5th, or even 7th overtone (overtones occur only at odd multiples). It's important to know whether the oscillator is operating in fundamental or overtone mode. An overtone is similar in concept to a harmonic, with the exception that crystal oscillation overtones are not exact integer multiples of the fundamental. Selection of overtone is based upon using the lowest possible overtone that will result in a crystal fundamental frequency below 30 MHz. The vendor calibrates a 3rd overtone crystal at the 3rd overtone, not the fundamental. For example, most crystal vendors will automatically give you a 3rd overtone 50 MHz crystal if you don't specify fundamental mode or an overtone mode. If you plug a 50 MHz 3rd overtone crystal into an oscillator that expects a fundamental-mode crystal, you are likely to have

an oscillator running at 50/3 or 16.666 MHz! If you don't know the frequency mode of your crystal, contact the designer or the manufacturer of the oscillator circuit.

The reason crystal vendors provide overtone crystals is that the quartz material becomes thinner and thinner as frequency increases. Starting at about 30 MHz, the quartz becomes so thin that it is hard to handle during the manufacturing process, and crystal vendors don't like to deal with thin crystals. One recent innovation in this area is the invention of inverted mesa crystals. Inverted mesa crystals can be manufactured with a thinner structure and thus can be reliably manufactured at higher fundamental-mode frequencies. This makes for less complex high-frequency oscillator designs and reduces component count by avoiding the need for external inductors/capacitors to induce the proper overtone oscillation mode from the crystal. Not all crystal vendors can provide inverted mesa technology; but, for the ones that can, they will be able to specify fundamental-mode crystals considerably higher than 30 MHz. Remember that an overtone-mode crystal cannot be used in a fundamental-mode oscillator, and vice versa. It may oscillate but not at the correct frequency.

Resonance Mode

Crystals have two modes of resonance: parallel and series. All crystals exhibit both resonance modes. The oscillator circuit is calibrated for one or the other, but not both. For applications requiring no tighter than 100ppm frequency accuracy, this spec is usually not an issue. However, if you are attempting to control frequency (or time) to within 100ppm, the resonance-mode specification becomes important. The difference to the crystal vendor is in which mode the crystal is calibrated during manufacturing. The crystal vendor sets up an oscillator circuit with the crystal in a customer-specified series resonance or parallel resonance and calibrates the crystal. Figure 2 shows crystal impedance behavior versus frequency as well as the relative location of each resonance mode.

Load Capacitance

Load capacitance is an important specification when using parallel resonant oscillation mode. Referring to Figure 2, it can be seen that crystal parallel resonance mode is always above the series resonance frequency and is characterized by inductive reactance. In parallel resonance oscillation mode, the crystal's inductance (motional inductance) is in parallel with the oscillator's load capacitance, thereby forming an LC tank circuit. This LC determines the oscillator frequency. If your oscillator uses parallel

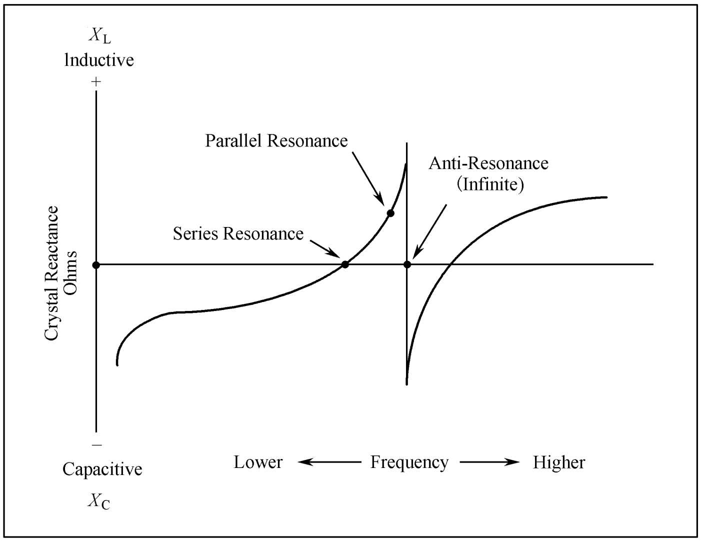


Figure 2 Crystal impedance versus frequency

resonance, the crystal vendor must know the load capacitance employed by the oscillator circuit. The load capacitance is simply the amount of external circuit capacitance in parallel with the crystal itself when it is placed in the oscillator circuit. The crystal vendor will then make sure that your crystal is calibrated at the factory using this same load capacitance. Vendors are flexible regarding load capacitance; ask them what range of load capacitance you can specify. Your oscillator should fall within the crystal vendor's acceptable range of load capacitance.

With a series resonant crystal, you can ignore the load capacitance specification, because the motional inductance and motional capacitance of the crystal are the only LC components that determine oscillation frequency.

Series Resistance

Series resistance is the effective resistive component in series with the LC model of the crystal itself (see Figure 1). Oscillator circuits can tolerate a certain degree of series resistance but not too much. A typical range is 25 Ohms to 100 Ohms for most crystals. The crystal vendor usually characterizes this resistance and specifies typical or maximum values for series resistance. Excessive crystal series resistance can lead to oscillator startup failure, so sufficient margin must be built into the oscillator design.

An exception is 32.768 kHz wristwatch crystals. Their series resistance can be in the tens of kilohms, so the oscillator circuit must be designed to accommodate this high

series resistance. Failure to address this will result in a 32.768 kHz oscillator that does not oscillate. Don't expect to use an oscillator designed for a 10MHz crystal with a 32.768 kHz crystal; it won't work.

Holder Capacitance

All crystals have small electrodes that connect the crystal to the package pins. The electrodes form a shunt capacitance in parallel with the crystal's LC model, as shown in Figure 1. Depending on the crystal's size and package, the holder capacitance can vary. Typical values range from 2 pF to 6 pF. Some oscillators will not tolerate excessive holder capacitance. This is particularly true at higher frequencies as the reactance of the holder capacitance decreases. Make sure the crystal vendor's holder capacitance is within the allowable range for your oscillator. As a general rule, minimize the holder capacitance (the smaller, the better).

Motional Inductance and Capacitance

Motional inductance and capacitance are specifications provided by the crystal vendor. They describe the L and C values that comprise the electrical LC model of the crystal. What is noteworthy about these values is the extreme ratio of L to C, which results in very large inductive and capacitive reactance values at the operating frequency. This is what gives a crystal its extraordinarily high "quality factor" also referred to as "Q" (Q being the ratio of energy stored to energy dissipated, also defined as the ratio of reactance to series resistance at the resonant frequency). For an LRC circuit, $Q = 1/R \times \text{sqrt.}\ (L/C)$ (the derivation of this is beyond the scope of this article). A high Q is desirable. Higher Q values mean less frequency shift for a change in oscillator load capacitance and less shift due to other external factors such as oscillator supply voltage. Depending on your application, your oscillator circuit may or may not require specification of motional inductance and capacitance.

Drive Level

The power dissipated in the crystal must be limited or the quartz crystal can actually fail due to excessive mechanical vibration. Crystal characteristics also change with drive level due to nonlinear behavior. Analyze the oscillator design to determine the power dissipated in the crystal. Power dissipated is the product of crystal current squared times crystal series resistance. For a parallel resonant oscillator, the crystal current equals the RMS voltage across the load capacitor divided by the load capacitor's

reactance at the oscillator frequency. For a series resonant crystal, the crystal current is the RMS voltage across the crystal divided by the crystal internal series resistance. The crystal manufacturer will specify maximum drive levels for a particular product line.

Expect crystal usage to continue to increase as crystals find their way into more and more products that use microcontrollers, digital signal processors, and data converters. Crystal technology is also moving forward, resulting in better performance and lower costs. Though at first glance, crystals may seem simple elements that one merely plugs into the circuit, an analysis of the actual circuit model and an understanding of key parameters show them in a different light and also simplify the process of designing them into your next application.

Questions:

1) Why is it important to understand the performance specifications of quartz crystals?

2) What is the frequency mode of crystals?

3) Can you tell the reasons why crystal vendors provide overtone crystals?

4) How many resonance modes do crystals have? What are they?

5) What is the trend of crystal technology?

Unit 4

Electronic Systems

Lesson 10 The Mobile Telephone System (Ⅰ)

Wireless telephones come in two basic varieties: cordless phones and mobile phones (sometimes called cell phones). Cordless phones are devices consisting of a base station and a handset sold as a set for use within the home. These are never used for networking. Mobile phones have gone through three distinct generations, with different technologies:

- Analog voice.
- Digital voice.
- Digital voice and data (Internet, e-mail, etc.).

The first mobile system was devised in the U. S. by AT&T[1] and mandated for the whole country by the FCC. As a result, the entire U. S. had a single (analog) system and a mobile phone purchased in California also worked in New York. In contrast, when mobile came to Europe, every country devised its own system, which resulted in a fiasco.

Europe learned from its mistake and when digital came around, the government-run PTTs[2] got together and standardized on a single system (GSM), so any European mobile phone will work anywhere in Europe. By then, the U. S. had decided that government should not be in the standardization business, so it left digital to the marketplace. This decision resulted in different equipment manufacturers producing different kinds of mobile phones. As a consequence, the U. S. now has two major incompatible digital mobile phone systems in operation (plus one minor one).

Despite an initial lead by the U. S., mobile phone ownership and usage in Europe is now far greater than in the U. S. Having a single system for all of Europe is part of the reason, but there is more. A second area where the U. S. and Europe differed is in the humble matter of phone numbers. In the U. S. mobile phones are mixed in with regular (fixed) telephones. Thus, there is no way for a caller to see if, say, (212) 234-5678 is a fixed telephone or a mobile phone. To keep people from getting nervous about using the telephone, the telephone companies decided to make the mobile phone owner pay for incoming calls. As a consequence, many people hesitated to buy a mobile phone for fear of running up a big bill by just receiving calls. In Europe, mobile phones have a special area code (analogous to 800 and 900 numbers) so they are instantly recognizable.

Consequently, the usual rule of "caller pays" also applies to mobile phones in Europe (except for international calls where costs are split).

A third issue that has had a large impact on adoption is the widespread use of prepaid mobile phones in Europe (up to 75% in some areas). These can be purchased in many stores with no more formality than buying a radio. You pay and you go. They are preloaded with, for example, 20 or 50 euro and can be recharged when the balance drops to zero. As a consequence, practically every teenager and many small children in Europe have (usually prepaid) mobile phones so their parents can locate them, without the danger of the child running up a huge bill. If the mobile phone is used only occasionally, its use is essentially free since there is no monthly charge or charge for incoming calls.

First-Generation Mobile Phones: Analog Voice

Mobile radiotelephones were used sporadically for maritime and military communication during the early decades of the 20th century. In 1946, the first system for car-based telephones was set up in St. Louis. This system used a single large transmitter on top of a tall building and had a single channel, used for both sending and receiving. To talk, the user had to push a button that enabled the transmitter and disabled the receiver. Such systems, known as push-to-talk systems, were installed in several cities beginning in the late 1950s. CB-radio, taxis, and police cars often use this technology.

In the 1960s, IMTS (Improved Mobile Telephone System) was installed. It, too, used a high-powered (200-watt[3]) transmitter, on top of a hill, but now had two frequencies, one for sending and one for receiving, so the push-to-talk button was no longer needed. Since all communication from the mobile telephones went inbound on a different channel than the outbound signals, the mobile users could not hear each other (unlike the push-to-talk system used in taxis). IMTS supported 23 channels spread out from 150 MHz to 450 MHz. Due to the small number of channels, users often had to wait a long time before getting a dial tone. Also, due to the large power of the hilltop transmitter, adjacent systems had to be several hundred kilometers apart to avoid interference. All in all, the limited capacity made the system impractical.

Advanced Mobile Phone System

All that changed with AMPS (Advanced Mobile Phone System), invented by Bell Labs and first installed in the United States in 1982. It was also used in England, where it was called TACS, and in Japan, where it was called MCS-L1. Many of its fundamental properties have been directly inherited by its digital successor, D-AMPS,

in order to achieve backward compatibility.

In all mobile phone systems, a geographic region is divided up into cells, which is why the devices are sometimes called cell phones. In AMPS, the cells are typically 10 to 20 km across; in digital systems, the cells are smaller. Each cell uses some set of frequencies not used by any of its neighbors. The key idea that gives cellular systems far more capacity than previous systems is the use of relatively small cells and the reuse of transmission frequencies in nearby (but not adjacent) cells. Whereas an IMTS system 100 km across can have one call on each frequency, an AMPS system might have 100 10-km cells in the same area and be able to have 10 to 15 calls on each frequency, in widely separated cells. Thus, the cellular design increases the system capacity by at least an order of magnitude[4], more as the cells get smaller. Furthermore, smaller cells mean that less power is needed, which leads to smaller and cheaper transmitters and handsets. Hand-held telephones put out 0.6 watts; transmitters in cars are 3 watts, the maximum allowed by the FCC.

The idea of frequency reuse is illustrated in Figure 10.1(a). The cells are normally roughly circular, but they are easier to model as hexagons. In Figure 10.1(a), the cells are all the same size. They are grouped in units of seven cells. Each letter indicates a group of frequencies. Notice that for each frequency set, there is a buffer about two cells wide where that frequency is not reused, providing for good separation and low interference.

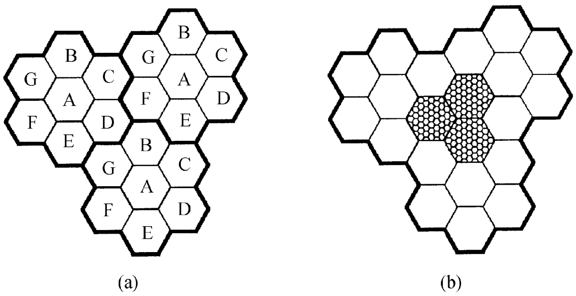


Figure 10.1 (a) Frequencies are not reused in adjacent cells.
(b) To add more users, smaller cells can be used

In an area where the number of users has grown to the point that the system is overloaded, the power is reduced, and the overloaded cells are split into smaller microcells to permit more frequency reuse, as shown in Figure 10.1(b). Telephone companies sometimes create temporary microcells[5], using portable towers with satellite links at sporting events, rock concerts, and other places where large numbers of mobile

users congregate for a few hours.

At the center of each cell is a base station to which all the telephones in the cell transmit. The base station consists of a computer and transmitter/receiver connected to an antenna. In a small system, all the base stations are connected to a single device called an MTSO (Mobile Telephone Switching Office) or MSC (Mobile Switching Center). In a larger one, several MTSOs may be needed, all of which are connected to a second-level MTSO, and so on. The MTSOs are essentially end offices as in the telephone system, and are, in fact, connected to at least one telephone system end office. The MTSOs communicate with the base stations, each other, and the PSTN using a packet-switching network.

New Words

mandate [ˈmændeit] *vt.* 委任

fiasco [fiˈæskəu] *n.* 惨败，大失败

humble [ˈhʌmbl] *adj.* 级别低的，位置低的

consequently [ˈkɔnsikwəntli] *adv.* 从而，因此

analogous [əˈnæləgəs] *adj.* 类似的

formality [fɔːˈmæliti] *n.* 手续，礼节，仪式

incoming [ˈinˌkʌmiŋ] *adj.* 输入的

install [inˈstɔːl] *vt.* 安装

high-powered [ˈhaiˈpauəd] *adj.* 大功率的

inbound [ˈinbaund] *adj.* 输入的

outbound [ˈautbaund] *adj.* 输出的

interference [ˌintəˈfiərəns] *n.* 干扰，干涉

capacity [kəˈpæsiti] *n.* 容量；电容

successor [səkˈsesə] *n.* 继承者，接任者；后续的事物

cell [sel] *n.* 细胞，蜂房，电池

cellular [ˈseljulə] *adj.* 蜂窝状的

circular [ˈsəːkjulə] *adj.* 圆形的，循环的

hexagon [ˈheksəgən] *n.* 六边形

buffer [ˈbʌfə] *n.* 缓冲器，缓冲区

second-level *adj.* 二级的

Phrases & Expressions

as a consequence 因此
for fear of 为了避免
run up 升起;积欠;匆匆制成
all in all 总而言之
order of magnitude 数量级
put out 放出,产生;消除;熄灭

Technical Terms

cordless [ˈkɔːdlis] *n.* 不用电线的
handset [ˈhændset] *n.* 电话听筒,手持机
watt [wɔt] *n.* 瓦特
overload [ˌəuvəˈləud] *vt.* 使超载,超过负荷 *n.* 超载,过载
antenna [ænˈtenə] *n.* 触角,天线
packet [ˈpækit] *n.* 封包,分组
microcell [ˈmaikrəsel] *n.* 微蜂窝
packet switching 分组交换
cordless phone 无绳电话
base station 基站
dial tone 拨号音
frequency reuse 频率重用
end office 端局
FCC *abbr.* Federal Communications Commission 联邦通信委员会
PTT *abbr.* Post Telephone and Telegraph Administration 邮电管理局
CB *abbr.* Citizens' Band 民用波段
MTSO *abbr.* Mobile Telephone Switching Office 移动电话交换局
MSC *abbr.* Mobile Switching Center 移动交换中心
PSTN *abbr.* Public Switched Telephone Network 公共交换电话网

Notes

1. 美国电话电报公司(American Telephone and Telegraph Corporation)曾是美国最

大的本地和长途电话公司，其前身是 1877 年创建的美国贝尔电话公司。

2. PTT 是对除美国、加拿大外的国家经营的公用电信部门的通称。
3. 瓦特(James Watt，1736－1819)是英国工程师、发明家。他对蒸汽机做出了基础性的改进，使之发展成现代的高压蒸汽机，并于 1769 年获得专利。
4. 数量级(order of magnitude)是指数量大小的级别，每个级别间存在固定比例。常用比例有 10、2、1000、1024 和 e，而十进制下的数量级是最常用的。如"两数相差 3 个数量级"就是说"一个数比另一个大 1000 倍"。
5. 微蜂窝(microcell)的发射功率较小(一般在 2 W 左右)，覆盖半径约为 100 m～ 1 km(如一幢建筑物或一个街区)。

Lesson 11　The Mobile Telephone System (Ⅱ)

Second-Generation Mobile Phones: Digital Voice

The first generation of mobile phones was analog; the second generation was digital. Just as there was no worldwide standardization during the first generation, there was also no standardization during the second, either. Four systems are in use now: D-AMPS, GSM, CDMA, and PDC. PDC is used only in Japan and is basically D-AMPS modified for backward compatibility with the first-generation Japanese analog system.

D-AMPS—The Digital Advanced Mobile Phone System

The second generation of the AMPS systems is D-AMPS and is fully digital. It is described in International Standard IS-54 and its successor IS-136. D-AMPS was carefully designed to coexist with AMPS so that both first- and second- generation mobile phones could operate simultaneously in the same cell. In particular, D-AMPS uses the same 30 kHz channels as AMPS and at the same frequencies so that one channel can be analog and the adjacent ones can be digital.

When D-AMPS was introduced as a service, a new frequency band was made available to handle the expected increased load. The upstream channels were in the 1850～1910 MHzrange, and the corresponding downstream channels were in the 1930～1990 MHzrange, again in pairs, as in AMPS. In this band, the waves are 16 cm long, so a standard 1/4 － wave antenna is only 4 cm long, leading to smaller phones. However, many D-AMPS phones can use both the 850-MHz and 1900-MHz bands to get a wider range of available channels.

On a D-AMPS mobile phone, the voice signal picked up by the microphone is

digitized and compressed using a model that is more sophisticated than the delta modulation and predictive encoding schemes. Compression takes into account detailed properties of the human vocal system to get the bandwidth from the standard 56-kbps PCM encoding to 8 kbps or less. The compression is done by a circuit called a vocoder [1]. The compression is done in the telephone, rather than in the base station or end office, to reduce the number of bits sent over the air link. With fixed telephony, there is no benefit to having compression done in the telephone, since reducing the traffic over the local loop does not increase system capacity at all.

With mobile telephony there is a huge gain from doing digitization and compression in the handset, so much so that in D-AMPS, three users can share a single frequency pair using time division multiplexing. Each frequency pair supports 25 frames/sec of 40 msec each. Each frame is divided into six time slots of 6.67 msec each, as illustrated in Figure 11.1(a) for the lowest frequency pair.

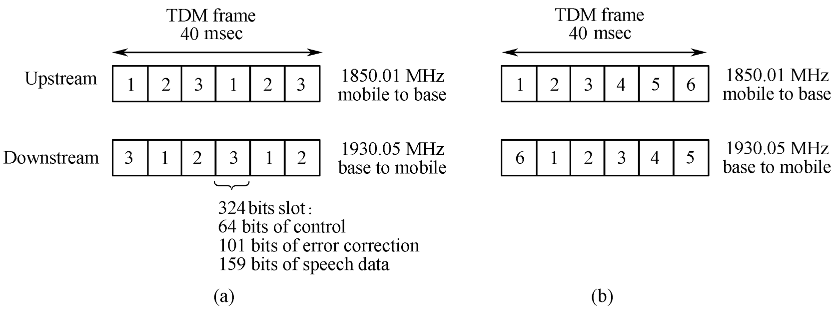


Figure 11.1 (a) A D-AMPS channel with 3 users. (b) A D-AMPS channel with 6 users

Each frame holds three users who take turns using the upstream and downstream links. During slot 1 of Figure 11.1(a), for example, user 1 may transmit to the base station and user 3 is receiving from the base station. Each slot is 324 bits long, of which 64 bits are used for guard times [2], synchronization, and control purposes, leaving 260 bits for the user payload. Of the payload bits, 101 are used for error correction over the noisy air link, so ultimately only 159 bits are left for compressed speech. With 50 slots/sec, the bandwidth available for compressed speech is just under 8 kbps, 1/7 of the standard PCM bandwidth.

Using better compression algorithms, it is possible to get the speech down to 4 kbps, in which case six users can be stuffed into a frame, as illustrated in Figure 11.1(b). From the operator's perspective, being able to squeeze three to six times as many D-AMPS users into the same spectrum as one AMPS user is a huge win

and explains much of the popularity of PCS.

GSM—The Global System for Mobile Communications

D-AMPS is widely used in the U. S. and (in modified form) in Japan. Virtually everywhere else in the world, a system called GSM (Global System for Mobile communications) is used, and it is even starting to be used in the U. S. on a limited scale. To a first approximation, GSM is similar to D-AMPS. Both are cellular systems. In both systems, frequency division multiplexing is used, with each mobile transmitting on one frequency and receiving on a higher frequency (80 MHz higher for D-AMPS, 55 MHz higher for GSM). Also in both systems, a single frequency pair is split by time-division multiplexing into time slots shared by multiple mobiles. However, the GSM channels are much wider than the D-AMPS channels (200 kHz versus 30 kHz) and hold relatively few additional users (8 versus 3), giving GSM a much higher data rate per user than D-AMPS.

Each frequency band is 200 kHz wide, as shown in Figure 11. 2. A GSM system has 124 pairs of simplex channels. Each simplex channel is 200 kHz wide and supports eight separate connections on it, using time division multiplexing. Each currently active station is assigned one time slot on one channel pair. Theoretically, 992 channels can be supported in each cell, but many of them are not available, to avoid frequency conflicts with neighboring cells. In Figure 11. 2, the eight shaded time slots all belong to the same connection, four of them in each direction. Transmitting and receiving does not happen in the same time slot because the GSM radios cannot transmit and receive at the same time and it takes time to switch from one to the other. If the mobile station assigned to 890. 4/935. 4 MHz and time slot 2 wanted to transmit to the base station, it would use the lower four shaded slots (and the ones following them in time), putting some data in each slot until all the data had been sent.

The TDM slots shown in Figure 11. 2 are part of a complex framing hierarchy. Each TDM slot has a specific structure, and groups of TDM slots form multiframes, also with a specific structure. A simplified version of this hierarchy is shown in Figure 11. 3. Here we can see that each TDM slot consists of a 148-bit data frame that occupies the channel for 577 μsec (including a 30-μsec guard time after each slot). Each data frame starts and ends with three 0 bits, for frame delineation purposes. It also contains two 57-bit Information fields, each one having a control bit that indicates whether the following Information field is for voice or data. Between the Information fields is a 26-bit Sync (training) field that is used by the receiver to synchronize to the sender's frame boundaries.

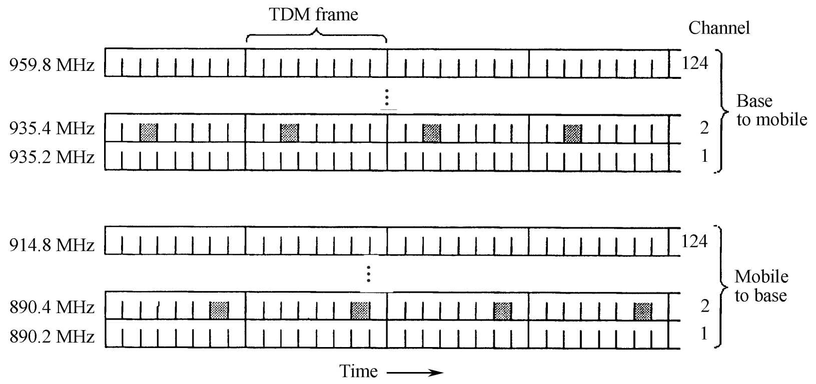


Figure 11.2 GSM uses 124 frequency channels, each of which uses an eight-slot TDM system

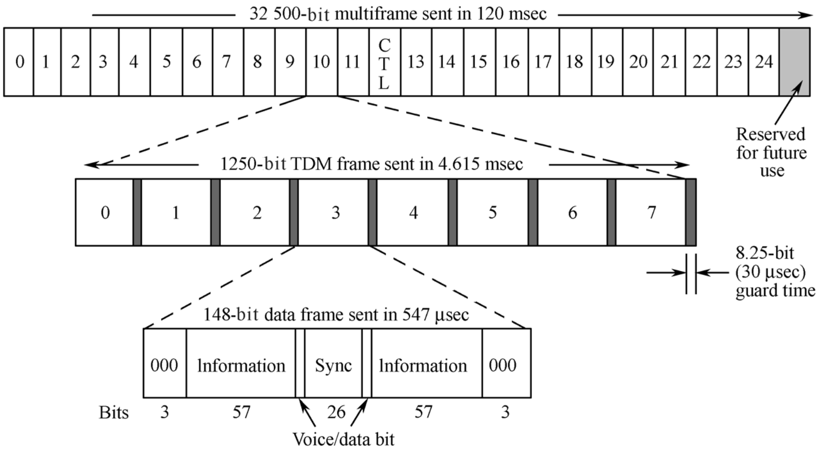


Figure 11.3 A portion of the GSM framing structure

A data frame is transmitted in 547 μsec, but a transmitter is only allowed to send one data frame every 4.615 msec, since it is sharing the channel with seven other stations. The gross rate of each channel is 270.833 kbps, divided among eight users. This gives 33.854 kbps gross, more than double D-AMPS' 50 times 324 bits per second for 16.2 kbps.

As can be seen from Figure 11.3, eight data frames make up a TDM frame and 26 TDM frames make up a 120-msec multiframe. Of the 26 TDM frames in a multiframe, slot 12 is used for control and slot 25 is reserved for future use, so only 24 are available for user traffic.

CDMA—Code Division Multiple Access

CDMA is completely different from AMPS, D-AMPS, and GSM. Instead of dividing the allowed frequency range into a few hundred narrow channels, CDMA allows each station to transmit over the entire frequency spectrum all the time. Multiple simultaneous transmissions are separated using coding theory. CDMA also relaxes the assumption that colliding frames are totally garbled. Instead, it assumes that multiple signals add linearly.

Let us consider an analogy: an airport lounge with many pairs of people conversing. TDM is comparable to all the people being in the middle of the room but taking turns speaking. FDM is comparable to the people being in widely separated clumps, each clump holding its own conversation at the same time as, but still independent of, the others. CDMA is comparable to everybody being in the middle of the room talking at once, but with each pair in a different language. The French-speaking couple just hones in on the French, rejecting everything that is not French as noise. Thus, the key to CDMA is to be able to extract the desired signal while rejecting everything else as random noise.

In CDMA, each bit time is subdivided into m short intervals called chips. Typically, there are 64 or 128 chips per bit, but in the example given below we will use 8 chips/bit for simplicity.

Each station is assigned a unique m-bit code called a chip sequence. To transmit a 1 bit, a station sends its chip sequence. To transmit a 0 bit, it sends the one's complement of its chip sequence. No other patterns are permitted. Thus, for $m = 8$, if station A is assigned the chip sequence 00011011, it sends a 1 bit by sending 00011011 and a 0 bit by sending 11100100.

Increasing the amount of information to be sent from b bits/sec to mb chips/sec can only be done if the bandwidth available is increased by a factor of m, making CDMA a form of spread spectrum communication (assuming no changes in the modulation or encoding techniques). If we have a 1-MHz band available for 100 stations, with FDM each one would have 10 kHz and could send at 10 kbps (assuming 1 bit per Hz). With CDMA, each station uses the full 1 MHz, so the chip rate is 1 megachip per second. With fewer than 100 chips per bit, the effective bandwidth per station is higher for CDMA than FDM, and the channel allocation problem is also solved.

In an ideal, noiseless CDMA system, the capacity (i. e. , number of stations) can be made arbitrarily large, just as the capacity of a noiseless Nyquist channel can be made arbitrarily large by using more and more bits per sample. In practice, physical

limitations reduce the capacity considerably. First, we have assumed that all the chips are synchronized in time. In reality, such synchronization is impossible. What can be done is that the sender and receiver synchronize by having the sender transmit a predefined chip sequence that is long enough for the receiver to lock onto. All the other (unsynchronized) transmissions are then seen as random noise. As one might expect, the longer the chip sequence, the higher the probability of detecting it correctly in the presence of noise. For extra reliability, the bit sequence can use an error-correcting code. Chip sequences never use error-correcting codes.

An implicit assumption in our discussion is that the power levels of all stations are the same as perceived by the receiver. CDMA is typically used for wireless systems with a fixed base station and many mobile stations at varying distances from it. The power levels received at the base station depend on how far away the transmitters are. A good heuristic here is for each mobile station to transmit to the base station at the inverse of the power level it receives from the base station. In other words, a mobile station receiving a weak signal from the base station will use more power than one getting a strong signal. The base station can also give explicit commands to the mobile stations to increase or decrease their transmission power.

We have also assumed that the receiver knows who the sender is. In principle, given enough computing capacity, the receiver can listen to all the senders at once by running the decoding algorithm for each of them in parallel. In real life, suffice it to say that this is easier said than done. CDMA also has many other complicating factors that have been glossed over here. Nevertheless, CDMA is a clever scheme that is being rapidly introduced for wireless mobile communication. It normally operates in a band of 1.25 MHz (versus 30 kHz for D-AMPS and 200 kHz for GSM), but it supports many more users in that band than either of the other systems. In practice, the bandwidth available to each user is at least as good as GSM and often much better.

New Words

coexist [kəuig'zist] *vi*. 共存

approximation [əˌprɔksi'meiʃn] *n*. 近似(值)

mobile ['məubail] *adj*. 运动的

delineation [diˌlini'eiʃn] *n*. 描绘,叙述

gross [grəus] *adj*. 总的

garble ['gɑːbl] *vt*. 混淆

collide [kə'laid] *vi*. 碰撞，抵触
analogy [ə'nælədʒi] *n*. 类似，类推
lounge [laundʒ] *n*. 休息室
converse [kən'vəːs] *vi*. 交谈
comparable ['kɔmpərəbl] *adj*. 可比较的，比得上的
clump [klʌmp] *n*. 块，团
hone [həun] *v*. 渴望，想念
subdivide ['sʌbdi'vaid] *v*. 再分，细分
pattern ['pætən] *n*. 图案，模式，结构
heuristic [hjuə'ristik] *adj*. 启发式的，探索的
inverse ['in'vəːs] *n*. 倒数，逆
computing [kəm'pjuːtiŋ] *n*. 计算，处理
real life *n*. 现实生活
complicate ['kɔmplikeit] *vt*. 使复杂，使难做，使恶化
clever ['klevə] *adj*. 精巧的，灵巧的，巧妙的
scheme [skiːm] *n*. 方案，计划

Phrases & Expressions

in practice 在实践中，实际上
in parallel 并行的，并行地
suffice it to say that 说……就足够了

Technical Terms

upstream [ˌʌp'striːm] *n*. 上行比特流
downstream ['daun'striːm] *n*. 下行比特流
vocoder ['vəuˌkəudə] *n*. 声码器
synchronization [ˌsiŋkrənai'zeiʃn] *n*. 同步
payload ['peiˌləud] *n*. 有效载荷
simplex ['simpleks] *n*. 单工，单向通信
multiframe ['mʌlti'freim] *n*. 复帧
delta modulation 增量调制
predictive encoding 预测编码
guard time 保护时间

error correction 纠错
frequency conflict 频率冲突
gross rate 总速率
coding theory 编码理论
random noise 随机噪声
one's complement 二进制反码
spread spectrum communication 扩频通信
weak signal 微弱信号
transmission power 发射功率
TDM *abbr*. Time Division Multiplexing 时分复用
PCM *abbr*. Pulse Code Modulation 脉冲编码调制
PCS *abbr*. Personal Communication Service 个人通信业务
FDM *abbr*. Frequency Division Multiplexing 频分复用

Notes

1. 声码器(vocoder)一词是由 voice 和 coder 组合而成的。
2. 保护时间(guard time)是为了防止 TDMA 系统中脉冲信号的重叠而提供的时间间隔。

Lesson 12　Personal Computer Systems

When you mention the word "technology", most people think about computers. Virtually every facet of our lives has some computerized component. The appliances in our homes have microprocessors built into them, as do our televisions. Even our cars have a computer. But the computer that everyone thinks of first is typically the personal computer, or PC.

A PC is a general-purpose tool built around a microprocessor. It has lots of different parts-memory, a hard disk, a modem, etc. —that work together. "General purpose" means that you can do many different things with a PC. You can use it to type documents, send e-mail, browse the Web and play games.

On the Inside

Let's take a look at the main components of a typical desktop computer.

• Central processing unit (CPU) — The microprocessor "brain" of the computer system is called the central processing unit. Everything that a computer does is overseen by the CPU.

• Memory — This is very fast storage used to hold data. It has to be fast because it connects directly to the microprocessor. There are several specific types of memory in a computer:

1) Random-access memory (RAM) — Used to temporarily store information that the computer is currently working with.
2) Read-only memory (ROM) — A permanent type of memory storage used by the computer for important data that does not change.
3) Basic input/output system (BIOS) — A type of ROM that is used by the computer to establish basic communication when the computer is first turned on.
4) Caching — The storing of frequently used data in extremely fast RAM that connects directly to the CPU.
5) Virtual memory — Space on a hard disk used to temporarily store data and swap it in and out of RAM as needed.

• Motherboard — This is the main circuit board that all of the other internal components connect to. The CPU and memory are usually on the motherboard. Other systems may be found directly on the motherboard or connected to it through a secondary connection. For example, a sound card can be built into the motherboard or connected through PCI.

• Power supply — An electrical transformer regulates the electricity used by the computer.

• Hard disk — This is large-capacity permanent storage used to hold information such as programs and documents.

• Operating system — This is the basic software that allows the user to interface with the computer.

• Integrated Drive Electronics (IDE) Controller — This is the primary interface for the hard drive, CD-ROM and floppy disk drive.

• Peripheral Component Interconnect (PCI) Bus — The most common way to connect additional components to the computer, PCI uses a series of slots on the

motherboard that PCI cards plug into.

- SCSI — the small computer system interface is a method of adding additional devices, such as hard drives or scanners, to the computer.
- AGP — Accelerated Graphics Port is a very high-speed connection used by the graphics card to interface with the computer.
- Sound card—This is used by the computer to record and play audio by converting analog sound into digital information and back again.
- Graphics card—This translates image data from the computer into a format that can be displayed by the monitor.

PC Connections

Input/Output

No matter how powerful the components inside your computer are, you need a way to interact with them. This interaction is called input/output (I/O). The most common types of I/O in PCs are:

- Monitor—The monitor is the primary device for displaying information from the computer.
- Keyboard—The keyboard is the primary device for entering information into the computer.
- Mouse—The mouse is the primary device for navigating and interacting with the computer.
- Removable storage — Removable storage devices allow you to add new information to your computer very easily, as well as save information that you want to carry to a different location.
    1) Floppy disk—The most common form of removable storage, floppy disks are extremely inexpensive and easy to save information to.
    2) CD-ROM —CD-ROM (compact disc, read-only memory) is a popular form of distribution of commercial software. Many systems now offer CD-R (recordable) and CD-RW (rewritable), which can also record.
    3) Flash memory — Based on a type of ROM called electrically erasable programmable read-only memory (EEPROM), Flash memory provides fast, permanent storage. CompactFlash, SmartMedia and PCMCIA cards are all types of Flash memory.
    4) DVD-ROM —DVD-ROM is similar to CD-ROM but is capable of holding

much more information.

Ports

- Parallel—This port is commonly used to connect a printer.
- Serial—This port is typically used to connect an external modem.
- Universal Serial Bus (USB) — Quickly becoming the most popular external connection, USB ports offer power and versatility and are incredibly easy to use.
- FireWire (IEEE 1394)—FireWire is a very popular method of connecting digital-video devices, such as camcorders or digital cameras, to your computer.

Internet/network

- Modem - This is the standard method of connecting to the Internet.
- Local area network (LAN) card—This is used by many computers, particularly those in an Ethernet office network, to connected to each other.
- Cable modem—Some people now use the cable-television system in their home to connect to the Internet.
- Digital Subscriber Line (DSL) modem — This is a high-speed connection that works over a standard telephone line.
- Very high bit-rate DSL (VDSL) modem — A newer variation of DSL, VDSL requires that your phone line have fiber-optic cables.

From Powerup to Shutdown

Now that you are familiar with the parts of a PC, let's see what happens in a typical computer session, from the moment you turn the computer on until you shut it down:

1. You press the "On" button on the computer and the monitor.
2. You see the BIOS software doing its thing, called the power-on self-test (POST). On many machines, the BIOS displays text describing such data as the amount of memory installed in your computer and the type of hard disk you have. During this boot sequence, the BIOS does a remarkable amount of work to get your computer ready to run.
   1) The BIOS determines whether the video card is operational. Most video cards have a miniature BIOS of their own that initializes the memory and graphics processor on the card. If they do not, there is usually video-driver information on another ROM on the motherboard that the BIOS can load.

2) The BIOS checks to see if this is a cold boot or a reboot. It does this by checking the value at memory address 0000:0472. A value of 1234h indicates a reboot, in which case the BIOS skips the rest of POST. Any other value is considered a cold boot.

3) If it is a cold boot, the BIOS verifies RAM by performing a read/write test of each memory address. It checks for a keyboard and a mouse. It looks for a PCI bus and, if it finds one, checks all the PCI cards. If the BIOS finds any errors during the POST, it notifies you with a series of beeps or a text message displayed on the screen. An error at this point is almost always a hardware problem.

4) The BIOS looks at the sequence of storage devices identified as boot devices in the CMOS Setup. "Boot" is short for "bootstrap", as in the old phrase "Lift yourself up by your bootstraps". [1] Boot refers to the process of launching the operating system. The BIOS tries to initiate the boot sequence from the first device using the bootstrap loader.

3. The bootstrap loader loads the operating system into memory and allows it to begin operation. It does this by setting up the divisions of memory that hold the operating system, user information and applications. The bootstrap loader then establishes the data structures that are used to communicate within and between the subsystems and applications of the computer. Finally, it turns control of the computer over to the operating system.

Once loaded, the operating system's tasks fall into six broad categories:

- Processor management—Breaking the tasks down into manageable chunks and prioritizing them before sending to the CPU
- Memory management—Coordinating the flow of data in and out of RAM and determining when virtual memory is necessary
- Device management—Providing an interface between each device connected to the computer, the CPU and applications
- Storage management—Directing where data will be stored permanently on hard drives and other forms of storage
- Application Interface—Providing a standard communication and data exchange between software programs and the computer
- User Interface—Providing a way for you to communicate and interact with the computer

When you choose the "Shut Down" option, the operating system closes all

programs that are currently active. If a program has unsaved information, you are given an opportunity to save it before closing the program. The operating system writes its current settings to a special configuration file so that it will boot up next time with the same settings. If the computer provides software control of power, then the operating system will completely turn off the computer when it finishes its own shut-down cycle. Otherwise, you will have to manually turn the power off.

New Words

facet [ˈfæsit] *n.* 面，方面
appliance [əˈplaiəns] *n.* 用具，器具
browse [brauz] *v.* 浏览
oversee [ˌəuvəˈsiː] *v.* 监视，检查
regulate [ˈregjuleit] *vt.* 控制，调节
graphics [ˈgræfiks] *n.* 图形
navigate [ˈnævigeit] *v.* 定位
cable [ˈkeibl] *n.* 电缆
session [ˈseʃn] *n.* 一段时间
sequence [ˈsiːkwəns] *n.* 顺序，序列
operational [ˌɔpəˈreiʃnl] *adj.* 操作的，运作的

Phrases & Expressions

on the inside 在内部，知道内情
now that 既然，由于
lift up 举起，使振奋
by one's (own) bootstraps 通过自己的努力

Technical Terms

modem [ˈməudəm] *n.* 调制解调器
e-mail [ˈiːmeil] *n.* 电子邮件
desktop [ˈdesktɔp] *n.* 台式电脑
motherboard [ˈmʌðəbɔːd] *n.* 主板，母板
transformer [trænsˈfɔːmə] *n.* 变压器，变换器

monitor [ˈmɔnitə] *n*. 监视器,监控器
mouse [maus] *n*. 鼠标器
camcorder [ˈkæmkɔːdə] *n*. 便携式摄像机
fiber [ˈfaibə] *n*. 光纤
shutdown [ˈʃʌtdaun] *n*. 关机
boot [buːt] *n*. 启动,引导,自举
reboot [riːˈbuːt] *n*. 重新启动
bootstrap [ˈbuːtstræp] *n*. 引导程序
powerup [ˌpauəˈʌp] *n*. 上电,加电
ethernet [ˈiːθənet] *n*. 以太网
virtual memory 虚拟内存
sound card 声卡
power supply 电源
operating system 操作系统
cable modem 线缆调制解调器
cable TV 有线电视
cold boot 冷启动
bootstrap loader 引导装入程序
processor management 处理器管理
memory management 内存管理
device management 设备管理
storage management 外存管理
application interface 应用程序接口
user interface 用户接口
PC *abbr*. Personal Computer 个人计算机
CPU *abbr*. Central Processing Unit 中央处理单元
PCI *abbr*. Peripheral Component Interconnect 周边元件互连接口
IDE *abbr*. Integrated Drive Electronics 集成驱动器电路
SCSI *abbr*. Small Computer System Interface 小型计算机系统接口
AGP *abbr*. Accelerated Graphic Port 加速图形接口
DVD *abbr*. Digital Video Disc 数字视盘
USB *abbr*. Universal Serial Bus 通用串行总线
LAN [læn] *abbr*. Local Area Network 局域网
POST *abbr*. Power-On Self Test 通电自检
DSL *abbr*. Digital Subscriber Line 数字用户线

VDSL *abbr*. Very high bit-rate DSL 超高数据率数字用户线

Notes

1. 此句可译为:boot 这个词就是 bootstrap 的略写形式,而这个 bootstrap 就是那句老话"Lift oneself up by one's bootstraps"中的那个 bootstrap。"lift oneself up by one's bootstraps"这个短语的意思是"靠自己的努力改善境遇"或者是"自立更生"。除了可以使用"lift"之外,还可以使用 raise 和 pull。
2. "集成驱动器电路"(IDE,读作[aid])是一种常见的硬盘、光驱接口标准。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) In a TDMA scheme several calls ______(共用一个频道). The scheme is useful for ______(数字化语音)or other digital data. Each call is allocated ______(若干时隙) based on its ______(数据率)within a frame for upstream ______ downstream. ______(除去) the user data, each time slot also carries other data for ______(同步), guard times, and control information.

The transmission ______ base station ______ mobile is done ______ time division multiplex (TDM) mode whereas ______ the upstream direction each mobile transmits in its own time slot. The overlap between different slots resulting ______ different propagation delay is prevented ______ using guard times and precise slot synchronization schemes.

The TDMA scheme is used along with the FDMA scheme because ______(在一个蜂窝内使用几个频道). The traffic in two directions is separated either by using two separate frequency channels ______ by alternating in time. The two schemes ______(被称做) frequency division duplex (FDD) and time division duplex (TDD), ______. The FDD scheme uses less bandwidth than TDD schemes use and does not require as precise synchronization of data flowing in two directions ______ in the TDD method. The latter, however, is useful when flexible bandwidth allocation is required for upstream and downstream traffic.

(2) The CDMA scheme is a direct sequence (DS), spread-spectrum method. It uses ______(线性调制) with wideband pseudonoise (PN) sequences to generate signals. These sequences, ______(也叫做)codes, spread the spectrum of the modulating signal ______ a large bandwidth, simultaneously reducing the ______(信号谱密度). Thus,

various CDMA signals occupy the same bandwidth and appear ______ noise to each other.

In the CDMA scheme, each user is assigned an individual code ______(在……时候) call initiation. This code is used both for spreading the signal at the time of transmission ______ despreading the signal at the time of reception. Cellular systems ______ (使用码分多址方式) use FDD, thus employing two frequency channels ______ forward and reverse links.

On forward-link a mobile transmits to all users synchronously and ______ preserves the orthogonality of various codes assigned to different users. The orthogonality, however, is not preserved between different components arriving from different paths ______ (在多径情形中). On reverse links each user transmits independently from other users ______(由于)their individual locations. Thus, the transmission on reverse link is ______(异步的)and the various signals are ______(不一定)orthogonal.

______(应当指出的是)these PN sequences are designed to be orthogonal to each other. In ______ words, the cross correlation between different code sequences is zero and thus the signal modulated with one code appears to be orthogonal to a receiver using a different code if the orthogonality is preserved during the transmission. This is the case on forward-link and ______ the absence of multipath the signal received by a mobile is not affected by ______(基站发送给其他移动用户的信号).

2. Translate the following passages into Chinese or English.

1) The available spectrum bandwidth is shared in a number of ways by various wireless radio links. The way in which this is done is referred to as a multiple access scheme. There are basically four principle schemes. These are frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and space division multiple access (SDMA).

2) The area served by mobile phone systems is divided into small areas known as cells. Each cell contains a base station that communicates with mobiles in the cell by transmitting and receiving signals on radio links. The transmission from the base station to a mobile is typically referred to as downstream, forwardlink, or downlink. The corresponding terms for the transmission from a mobile to the base are upstream, reverse-link, and uplink. Each base station is associated with a mobile switching center (MSC) that connects calls to and from the base to mobiles in other cells and the public switched telephone network.

3) A base station communicates with mobiles using two types of radio channels,

control channels to carry control information and traffic channels to carry messages.

4) In an FDMA scheme the available spectrum is divided into a number of frequency channels of certain bandwidth and individual calls use different frequency channels. All first-generation cellular systems use this scheme.

5) In TDMA and FDMA systems a frequency channel used in a cell is not used in adjacent cells to prevent co-channel interference. In a CDMA system it is possible to use the same frequency channel in adjacent cells and thus increase the system capacity.

6)“信道”这个术语通常是指频分多址系统中的一个频率,时分多址系统中的一个时隙,码分多址系统中的一个代码或混合系统中频率、时隙、代码的某种组合。

7)在频分多址和时分多址系统中,一旦信道(频率或时隙)分配给用户,该信道在非活动期间也不能被利用。

8)“蜂窝”概念是由贝尔实验室发明的。1983 年 10 月,在芝加哥出现了第一个商用模拟语音系统。

9)第三代系统的目的在于为用户提供一个支持语音、数据、多媒体和视频服务的无缝网络,而不论用户在网络中的什么位置。

10)在第三代通信系统中,卫星将在提供全球覆盖方面发挥重要作用。

Reading Materials

Passage 1 The Future of Computing

Silicon microprocessors have been the heart of the computing world for more than 40 years. In that time, microprocessor manufacturers have crammed more and more electronic devices onto microprocessors. In accordance with Moore's Law, the number of electronic devices put on a microprocessor has doubled every 18 months. Moore's Law is named after Intel founder Gordon Moore, who predicted in 1965 that microprocessors would double in complexity every two years. Many have predicted that Moore's Law will soon reach its end because of the physical limitations of silicon microprocessors.

The current process used to pack more and more transistors onto a chip is called deep-ultraviolet lithography (DUVL), which is a photography-like technique that focuses light through lenses to carve circuit patterns on silicon wafers. DUVL will begin to reach its limit around 2005. At that time, chipmakers will have to look to other technologies to cram more transistors onto silicon to create more powerful chips. Many are already looking at extreme-ultraviolet lithography (EUVL) as a way to extend the life of silicon at least until the end of the decade. EUVL uses mirrors instead of lenses to focus the light, which allows light with shorter wavelengths to accurately focus on the silicon wafer.

Beyond EUVL, researchers have been looking at alternatives to the traditional microprocessor design. Two of the more interesting emerging technologies are DNA computers and quantum computers.

DNA computers have the potential to take computing to new levels, picking up where Moore's Law leaves off. There are several advantages to using DNA instead of silicon:

As long as there are cellular organisms, there will be a supply of DNA. The large supply of DNA makes it a cheap resource.

Unlike traditional microprocessors, which are made using toxic materials, DNA biochips can be made cleanly.

DNA computers are many times smaller than today's computers.

DNA's key advantage is that it will make computers smaller, while at the same

time increasing storage capacity, than any computer that has come before. One pound of DNA has the capacity to store more information than all the electronic computers ever built. The computing power of a teardrop-sized DNA computer, using the DNA logic gates, will be more powerful than the world's most powerful supercomputer. More than 10-trillion DNA molecules can fit into an area no larger than 1 cubic centimeter (.06 $inch^3$). With this small amount of DNA, a computer would be able to hold 10 terabytes (TB) of data and perform 10-trillion calculations at a time. By adding more DNA, more calculations could be performed.

Unlike conventional computers, DNA computers could perform calculations simultaneously. Conventional computers operate linearly, taking on tasks one at a time. It is parallel computing that will allow DNA to solve complex mathematical problems in hours—problems that might take electrical computers hundreds of years to complete.

Today's computers work by manipulating bits that exist in one of two states: 0 or 1. Quantum computers aren't limited to two states; they encode information as quantum bits, or qubits. A qubit can be a 1 or a 0, or it can exist in a superposition that is simultaneously 1 and 0 or somewhere in between. Qubits represent atoms that are working together to serve as computer memory and a microprocessor. Because a quantum computer can contain these multiple states simultaneously, it has the potential to be millions of times more powerful than today's most powerful supercomputers. A 30-qubit quantum computer would equal the processing power of a conventional computer capable of running at 10 teraops, or trillions of operations per second. Today's fastest supercomputers have achieved speeds of about 2 teraops.

Already we are seeing powerful computers in non-desktop roles. Laptop computers and personal digital assistants (PDAs) have taken computing out of the office. Wearable computers built into our clothing and jewelry will be with us everywhere we go. Our files will follow us while our computer provides constant feedback about our environment. Voice- and handwriting-recognition software will allow us to interface with our computers without using a mouse or keyboard. Magnetic RAM and other innovations will soon provide our PC with the same instant-on accessibility that our TV and radio have.

One thing is an absolute certainty: The PC will evolve. It will get faster. It will have more capacity. And it will continue to be an integral part of our lives.

Questions:

1) What does Moore's Law say?

2) What does ***DUVL*** stand for? And what about ***EUVL***?

3) What advantages does a DNA computer have over its silicon counterpart?

4) How does a quantum computer differ from a conventional computer?

5) What do you think of the future of computing?

Passage 2 Satellite-Based Mobile Communications

Mobile satellite communications began in 1976 with the launch by COMSAT of the MARISAT satellites to provide communications to ships at sea. The International Maritime Satellite Organization (INMARSAT) was subsequently formed in 1979, and that organization now provides mobile satellite communications services to aircraft and land-based terminals. A number of national mobile satellite communications systems also serve the United States, Canada, Australia, and Japan with many more planned.

The spectacular growth of terrestrial mobile communications systems has provided a catalyst for efforts to provide global mobile communications through the use of mobile satellite communications systems in low-, medium-, and geostationary-Earth orbit.

Until now, second-generation terrestrial and satellite mobile communications systems have existed as two independent environments. However, these environments are beginning to combine to form a third-generation global mobile communications system in which terrestrial and satellite systems have complementary instead of independent roles and form a single universal integrated system.

A Brief History of Satellite Communications

In an article in Wireless World in 1945, Arthur C. Clarke proposed the idea of placing satellites in geostationary orbit around Earth such that three equally spaced satellites could provide worldwide coverage. However, it was not until 1957 that the Soviet Union launched the first satellite Sputnik 1, which was followed in early 1958 by the U. S. Army's Explorer 1. Both Sputnik and Explorer transmitted telemetry information.

The first communications satellite, the Signal Communicating Orbit Repeater Experiment (SCORE), was launched in 1958 by the U. S. Air Force. SCORE was a delayed-repeater satellite, which received signals from Earth at 150 MHz and stored them on tape for later retransmission. A further experimental communication satellite, Echo 1, was launched on August 12, 1960 and placed into inclined orbit at about 1500 km

above Earth. Echo 1 was an aluminized plastic balloon with a diameter of 30 m and a weight of 75.3 kg. Echo 1 successfully demonstrated the first two-way voice communications by satellite.

On October 4, 1960, the U. S. Department of Defense launched Courier into an elliptical orbit between 956 and 1240 km, with a period of 107 min. Although Courier lasted only 17 days, it was used for real-time voice, data, and facsimile transmission. The satellite also had five tape recorders onboard; four were used for delayed repetition of digital information, and the other for delayed repetition of analog messages.

Direct-repeated satellite transmission began with the launch of Telstar Ⅰ on July 10, 1962. Telstar Ⅰ was an 87-cm, 80-kg sphere placed in low-Earth orbit between 960 and 6140 km, with an orbital period of 158 min. Telstar Ⅰ was the first satellite to be able to transmit and receive simultaneously and was used for experimental telephone, image, and television transmission. However, on February 21, 1963, Telstar Ⅰ suffered damage caused by the newly discovered Van Allen belts.

Telstar Ⅱ was made more radiation resistant and was launched on May 7, 1963. Telstar Ⅱ was a straight repeater with a 6.5-GHz uplink and a 4.1-GHz downlink. The satellite power amplifier used a specially developed 2-W traveling wave tube. Along with its other capabilities, the broadband amplifier was able to relay color TV transmissions. The first successful trans-Atlantic transmission of video was accomplished with Telstar Ⅱ, which also incorporated radiation measurements and experiments that exposed semiconductor components to space radiation.

The first satellites placed in geostationary orbit were the synchronous communication (SYNCOM) satellites launched by NASA in 1963. SYNCOM Ⅰ failed on injection into orbit. However, SYNCOM Ⅱ was successfully launched on July 26, 1964 and provided telephone, teletype, and facsimile transmission. SYNCOM Ⅲ was launched on August 19, 1964 and transmitted TV pictures from the Tokyo Olympics. The International Telecommunications by Satellite (INTELSAT) consortium was founded in July 1964 with the charter to design, construct, establish, and maintain the operation of a global commercial communications system on a nondiscriminatory basis. The INTELSAT network started with the launch on April 6, 1965, of INTELSAT Ⅰ, also called Early Bird. On June 28, 1965, INTELSAT Ⅰ began providing 240 commercial international telephone channels as well as TV transmission between the United States and Europe.

In 1979, INMARSAT established a third global system. In 1995, the INMARSAT name was changed to the International Mobile Satellite Organization to reflect the fact

that the organization had evolved to become the only provider of global mobile satellite communications at sea, in the air, and on the land.

Early telecommunication satellites were mainly used for long-distance continental and intercontinental broadband, narrowband, and TV transmission. With the advent of broadband optical fiber transmission, satellite services shifted focus to TV distribution, and to point-to-multipoint and very small aperture terminal (VSAT) applications. Satellite transmission is currently undergoing further significant growth with the introduction of mobile satellite systems for personal communications and fixed satellite systems for broadband data transmission.

Types of Telecommunications Satellite Services

Because satellite communications cover the whole range of voice, data, and video transmission, telecommunication satellite services are normally classified into three types:

- *Fixed satellite service* (*FSS*) networks are mainly intended for long-distance operation of telecommunications networks. FSS satellites are employed to relay signals between large, complex, and expensive Earth stations, which are connected to the terrestrial telecommunications network.
- *Direct-broadcast satellite service* (*DBS*) networks transmit broadcast and TV signals from a large central Earth station, via a satellite to receive-only Earth stations. DBS receive stations either are distribution heads for cable TV or are located in homes for direct-to-home transmission.
- *Mobile satellite services* (*MSS*) networks are relayed via satellite between large fixed Earth stations and small mobile terminals fitted to a ship, an aircraft, or a vehicle. Increasingly, MSS networks are formed to relay communications to portable handheld terminals.

In 1996, the ITU defined the Global Mobile Personal Communications by Satellite (GMPCS) as comprising the following systems:

- *Geostationary Earth Orbit* (*GEO*) *MSS* are for voice and low-speed data mobile personal communications services.
- *Non-GEO* (*NGEO*) *MSS* are for narrowband mobile personal communications services excluding voice - because these are invariably based on low-Earth orbit (LEO) satellites, they are also called Little-LEO.
- *NGEO MSS* for narrowband mobile personal communications include voice, operating in LEO, medium-Earth orbit (MEO), or highly elliptical orbit (HEO) -also

called Big-LEO.

• *GEO and NGEO FSS* offer fixed and transportable multimedia broadband services—also called Super-LEO.

Questions:

1) Who proposed the idea of satellite communication?

2) When was the first communications satellite launched?

3) Which satellite first accomplished the trans-Atlantic transmission of video?

4) What are telecommunication satellites mainly used for today?

5) How many types of services can telecommunication satellites provide?

Passage 3 The Global Positioning System(GPS)

What is GPS?

GPS, which stands for Global Positioning System, is the only system today able to show you your exact position on the Earth anytime, in any weather, anywhere. GPS satellites, 24 in all, orbit 11,000 nautical miles above the Earth. They are continuously monitored by ground stations located worldwide. The satellites transmit signals that can be detected by anyone with a GPS receiver. Using the receiver, you can determine your location with great precision.

What is Navigation?

Since prehistoric times, people have been trying to figure out a reliable way to tell where they are and to help guide them to where they are going. Cavemen probably used stones and twigs to mark a trail when they set out hunting for food. The earliest mariners followed the coast closely to keep from getting lost. When navigators first sailed into the open ocean, they discovered they could chart their course by following the stars. The ancient Phoenicians used the North Star to journey from Egypt and Crete. According to Homer, the goddess Athena told Odysseus to"keep the Great Bear on his left"during his travels from Calypso's Island. Unfortunately, the stars are only visible at night-and only on clear nights.

The next major developments in the quest for the perfect method of navigation were the magnetic compass and the sextant. The needle of a compass always points north, so it is always possible to know in what direction you are going. The sextant uses

adjustable mirrors to measure the exact angle of the stars, moon, and sun above the horizon. However, in the early days of its use, it was only possible to determine latitude (the location on the Earth measured north or south from the equator) from the sextant observations. Sailors were still unable to determine their longitude (the location on the Earth measured east or west). This was such a serious problem that in the 17th century, the British formed a special Board of Longitude consisting of well-known scientists.

This group offered £20,000, equal to about a million of today's dollars, to anybody who could find a way to determine a ship's longitude within 30 nautical miles. The generous offer paid off. In 1761, a cabinetmaker named John Harrison developed a shipboard timepiece called a chronometer, which lost or gained only about one second a day - incredibly accurate for the time. For the next two centuries, sextants and chronometers were used in combination to provide latitude and longitude information. In the early 20th century several radio-based navigation systems were developed, which were used widely during World War Ⅱ. Both allied and enemy ships and airplanes used ground-based radio-navigation systems as the technology advanced.

A few ground-based radio-navigation systems are still in use today. One drawback of using radio waves generated on the ground is that you must choose between a system that is very accurate but doesn't cover a wide area, or one that covers a wide area but is not very accurate. High-frequency radio waves (like UHF TV) can provide accurate position location but can only be picked up in a small, localized area. Lower frequency radio waves (like AM radio) can cover a larger area, but are not a good yardstick to tell you exactly where you are.

Scientists, therefore, decided that the only way to provide coverage for the entire world was to place high-frequency radio transmitters in space. A transmitter high above the Earth sending a high-frequency radio wave with a special coded signal can cover a large area and still overcome much of the "noise" encountered on the way to the ground. This is one of the main principles behind the GPS system.

GPS Elements

GPS has 3 parts: the space segment, the user segment, and the control segment. The space segment consists of 24 satellites, each in its own orbit 11,000 nautical miles above the Earth. The user segment consists of receivers, which you can hold in your hand or mount in your car. The control segment consists of ground stations (five of them, located around the world) that make sure the satellites are working properly.

One trip around the Earth in space equals one orbit. The GPS satellites each take 12 hours to orbit the Earth. Each satellite is equipped with an accurate clock to let it broadcast signals coupled with a precise time message. The ground unit receives the satellite signal, which travels at the speed of light. Even at this speed, the signal takes a measurable amount of time to reach the receiver.

The difference between the time the signal is sent and the time it is received, multiplied by the speed of light, enables the receiver to calculate the distance to the satellite. To measure precise latitude, longitude, and altitude, the receiver measures the time it took for the signals from four separate satellites to get to the receiver.

The GPS system can tell you your location anywhere on or above the Earth to within about 300 feet. Even greater accuracy, usually within less than three feet, can be obtained with corrections calculated by a GPS receiver at a known fixed location.

Satellites in Space

As we've said, the complete GPS space system includes 24 satellites, 11,000 nautical miles above the Earth, which take 12 hours to go around the Earth once (one orbit). They are positioned so that we can receive signals from six of them nearly 100 percent of the time at any point on Earth. You need that many signals to get the best position information.

Satellites are equipped with very precise clocks that keep accurate time to within three nanoseconds—that's 0.000000003, or three billionths, of a second. This precision timing is important because the receiver must know exactly how long it takes for its signal to get to each satellite and return. By knowing the exact amount of time the signal has taken to get back from each satellite, it can calculate its position.

The first GPS satellite was launched in 1978. The first 10 satellites were developmental satellites, called Block Ⅰ. From 1989 to 1993, 23 production satellites, called Block Ⅱ, were launched. The launch of the 24th satellite in 1994 completed the system.

Ground Control Stations

The GPS control, or ground, segment consists of unmanned monitor stations located around the world (Hawaii and Kwajalein in the Pacific Ocean; Diego Garcia in the Indian Ocean; Ascension Island in the Atlantic Ocean; and Colorado Springs, Colorado); a master ground station at Falcon Air Force Base in Colorado Springs, Colorado; and four large ground antenna stations that broadcast signals to the satellites.

The stations also track and monitor the GPS satellites.

Receivers

GPS receivers can be hand carried or installed on aircraft, ships, tanks, submarines, cars, and trucks. These receivers detect, decode, and process GPS satellite signals. More than 100 different receiver models are already in use. The typical hand-held receiver is about the size of a cellular telephone, and the newer models are even smaller.

How GPS Works

So you can more easily understand some of the scientific principles that make GPS work, let's discuss the basic features of the system. The principle behind GPS is the measurement of distance (or "range") between the receiver and the satellites. The satellites also tell us exactly where they are in their orbits. It works something like this: If we know our exact distance from a satellite in space, we know we are somewhere on the surface of an imaginary sphere with radius equal to the distance to the satellite radius. If we know our exact distance from two satellites, we know that we are located somewhere on the line where the two spheres intersect. And, if we take a third measurement, there are only two possible points where we could be located. One of these is usually impossible, and the GPS receivers have mathematical methods of eliminating the impossible location.

GPS Uses in Everyday Life

The GPS system was developed to meet military needs of the Department of Defense, but new ways to use its capabilities are continually being found. As you have read, the system has been used in aircraft and ships, but there are many other ways to benefit from GPS. We'll mention just a few.

During construction of the tunnel under the English Channel, British and French crews started digging from opposite ends: one from Dover, England, one from Calais, France. They relied on GPS receivers outside the tunnel to check their positions along the way and to make sure they met exactly in the middle. Otherwise, the tunnel might have been crooked.

Remember the example of the car with a video display in the dashboard? Vehicle tracking is one of the fastest-growing GPS applications. GPS-equipped fleet vehicles, public transportation systems, delivery trucks, and courier services use receivers to

monitor their locations at all times.

GPS is also helping to save lives. Many police, fire, and emergency medical service units are using GPS receivers to determine the police car, fire truck, or ambulance nearest to an emergency, enabling the quickest possible response in life-or-death situations.

Automobile manufacturers are offering moving-map displays guided by GPS receivers as an option on new vehicles. The displays can be removed and taken into a home to plan a trip. Several Florida rental car companies are demonstrating GPS-equipped vehicles that give directions to drivers on display screens and through synthesized voice instructions. No more getting lost on the way to Disney World!

Mapping and surveying companies use GPS extensively. In the field of wildlife management, endangered species such as Montana elk and Mojave Desert tortoises are being fitted with GPS receivers and tiny transmitters to help determine population distribution patterns and possible sources of disease.

GPS-equipped balloons are monitoring holes in the ozone layer over the Polar Regions, and air quality is being monitored using GPS receivers. Buoys tracking major oil spills transmit data using GPS. Archaeologists and explorers are using the system. Anyone equipped with a GPS receiver can use it as a reference point to locate or find another location. With a basic knowledge of math and science, plus a hand-held GPS receiver, you could be an instant hero if you and friends got temporarily lost on a camping trip.

The future of GPS is as unlimited as your imagination. New applications will continue to be created as technology evolves. The GPS satellites, like handmade stars in the sky, will be guiding you well into the future.

Questions:

1) When did the first radio-based navigation system appear?

2) How is GPS constructed?

3) Why is precision timing important in GPS?

4) When was GPS completed?

5) Can you give some examples of GPS application in your daily life?

Unit 5

Modern Digital Design

Lesson 13　Overview of Modern Digital Design

Electronic circuit design has traditionally fallen into two main areas: analogue and digital. These subjects are usually taught separately, and electronics engineers tend to specialize in one area. Within these two groupings there are further specializations, such as radio frequency analogue design; digital integrated circuit design; and, where the two domains meet, mixed-signal design. In addition, of course, software engineering plays an increasingly important role in embedded systems.

Digital electronics is ever more significant in consumer goods. Cars have sophisticated control systems. Many homes now have personal computers. Products that used to be thought of as analogue, such as radio, television and telephones, are all becoming digital. Digital compact discs have almost entirely replaced analogue LPs for recorded audio. With these changes, the lifetimes of products have lessened. In a period of less than a year, new models will probably have replaced all the digital electronic products in your local store.

Design automation

To keep pace with this rapid change, electronics products have to be designed extremely quickly. Analogue design is still a specialized (and well-paid) profession. Digital design has become very dependent on computer-aided design (CAD)-also known as design automation (DA) or electronic design automation (EDA). The EDA tools allow two tasks to be performed: synthesis, in other words the translation of a specification into an actual implementation of the design; and simulation, in which the specification or the detailed implementation can be exercised in order to verify correct operation [1].

Synthesis and simulation EDA tools require that the design be transferred from the designer's imagination into the tools themselves. This can be done by drawing a diagram of the design using a graphical package. This is known as schematic capture. Alternatively, the design can be represented in a textual form, much like a software program. Textual descriptions of digital hardware can be written in a modified programming language, such as C, or in a hardware description language (HDL). Over the past thirty years or so, a number of HDLs have been designed. Two HDLs are in common usage today: Verilog and VHDL (VHSIC Hardware Description Language, where VHSIC stands for Very High Speed

Integrated Circuit). Standard HDLs are important because they can be used by different CAD tools from different tool vendors. In the days before Verilog and VHDL, every tool had its own HDL, requiring laborious translation between HDLs, for example to verify the output from a synthesis tool with another vendor's simulator.

Logic gates

The basic building blocks of digital circuits are gates. A gate is an electronic component with a number of inputs and, generally, a single output. The inputs and the outputs are normally in one of two states: logic 0 or logic 1. These logic values are represented by voltages (for instance, 0 V for logic 0 and 3.3 V for logic 1) or currents. The gate itself performs a logical operation using all of its inputs to generate the output. Ultimately, of course, digital gates are really analogue components, but for simplicity we tend to ignore their analogue nature.

It is possible to buy a single integrated circuit containing, say, four identical gates, as shown in Figure 13.1. (Note that two of the connections are for the positive and negative power supplies to the device. These connections are not normally shown in logic diagrams.) A digital system could be built by connecting hundreds of such devices together - indeed many systems have been designed in that way. Although the individual integrated circuits might cost as little as 10 cents each, the cost of designing the printed circuit board for such a system and the cost of assembling the board are very significant and this design style is no longer cost-effective.

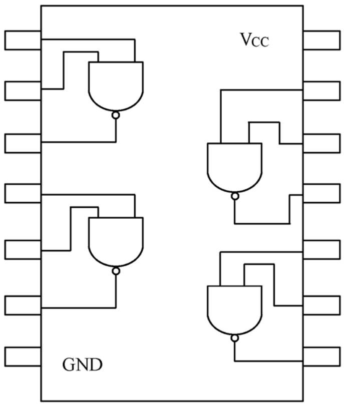


Figure 13.1 A small-scale integrated circuit

Much more complicated functions are available as mass-produced integrated circuits, ranging from flip-flops through to microprocessors. With increasing complexity comes

flexibility—a microprocessor can be programmed to perform a near-infinite variety of tasks. Digital system design therefore consists, in part, of taking standard components and connecting them together. Inevitably, however, some aspect of the functionality will not be available as a standard device. The designer is then left with the choice of implementing this functionality from discrete gates or of designing a specialized integrated circuit to perform that task. While this latter task may appear daunting, it should be remembered that the cost of a system will depend to a great extent not on the cost of the individual components but on the cost of connecting those components together.

ASICs and FPGAs

The design of a high-performance, full-custom integrated circuit (IC) is, of course, a difficult task. In full-custom IC design, everything, down to and including individual transistors may be designed (although libraries of parts are, of course, used). For many years, however, it has been possible to build semi-custom integrated circuits using gate arrays. A gate array, as its name suggests, is an integrated circuit on which an array of logic gates has been created. The design of an application-specific integrated circuit (ASIC) using a gate array therefore involves the definition of how the gates in the array should be connected. In practical terms, this means that one or two layers of metal interconnect must be designed. Since an integrated circuit requires seven or more processing stages, all the processing steps other than the final metalization can be completed in advance. Because the uncommitted gate arrays can be produced in volume, the cost of each device is relatively small.

The term ASIC is often applied to full-custom and semi-custom integrated circuits. Another class of integrated circuit is that of programmable logic. The earliest programmable logic devices (PLDs) were programmable logic arrays (PLAs). Like gate arrays, these consist of arrays of uncommitted logic, but unlike mask-programmable gate arrays, the configuration of the array is determined by applying a large (usually negative) voltage to individual connections. The general structure of a PLA is shown in Figure 13.2. The PLA has a number of inputs (A, B, C) and outputs (X, Y, Z), an AND-plane and an OR-plane. Connections between the inputs and the product terms (P, Q, R, S) and between the product terms and outputs are shown; the remaining connections have been removed as part of the programming procedure. Some PLAs may be reprogrammed electrically or by restoring the connections by exposing the device to ultraviolet light. PALs (Programmable Array Logic) extend the idea of PLAs to include up to 12 flip-flops. In recent years, programmable devices have become much more complex and include CPLDs (complex PLDs) and FPGAs (Field Programmable Gate Arrays).

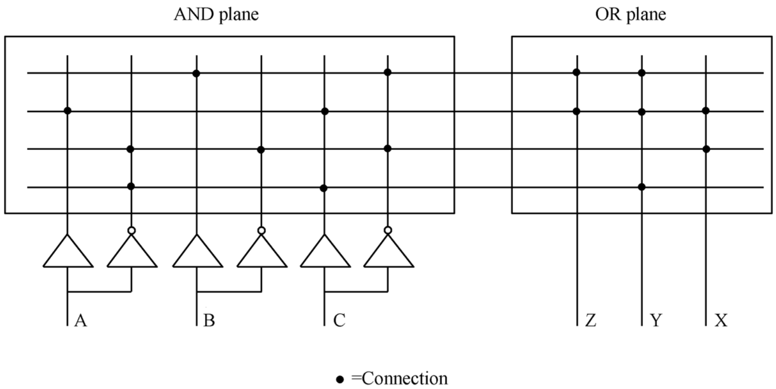


Figure 13.2 PLA structure

Design flow

Most digital systems are sequential, that is they have states, and the outputs depend on the present state. Some early designs of computer were asynchronous; in other words, the transition to a new state happened as soon as inputs had stabilized. For many years, digital systems have tended to be synchronous. In a synchronous system, the change of state is triggered by one or more clock signals. In order to design reliable systems, formal design methodologies have been defined. The design of a (synchronous sequential digital system using discrete gates would therefore proceed as follows.

1. Write a specification.

2. If necessary, partition the design into smaller parts and write a specification for each part.

3. From the specification draw a state machine chart. This shows each state of the system and the input conditions that cause a change of state, together with the outputs in each state.

4. Minimize the number of states. This is optional and may not be useful in all cases.

5. Assign Boolean[2] variables to represent each state.

6. Derive next state and output logic.

7. Optimize the next state and output logic to minimize the number of gates needed.

8. Choose a suitable placement for the gates in terms of which gates share integrated circuits and in terms of where each integrated circuit is placed on the printed circuit board[3].

9. Design the routing between the integrated circuits.

In general, steps 1 and 2 cannot be avoided. This is where the creativity of the

designer is needed. Most books on digital design concentrate on steps 3 to 7. Steps 8 and 9 can be performed manually, but placement and routing was one of the first tasks to be successfully automated. It is possible to simulate the design at different stages if it is converted into a computer-readable form. Typically, in order to perform the placement and routing, a schematic capture program would be used at around step 7, such that the gate-level structure of the circuit would be entered. This schematic could be converted to a form suitable for a logic simulator. After step 9 had been completed, the structure of the circuit, including any delays generated by the resistance and capacitance of the interconnect could be extracted and again simulated.

The implementation of digital designs on ASICs or FPGAs therefore involves the configuration of connections between predefined logic blocks. As noted, we cannot avoid steps 1 and 2, above and steps 8 and 9 can be done automatically. The use of an HDL, in the case of this book VHDL, means that the design can be entered into a CAD system and simulated at step 3 or 4, rather than step 7. So-called register transfer level (RTL) synthesis tools automate steps 6 and 7. Step 4 still has to be done by hand. Step 5 can be automated, but now the consequences of a particular state assignment can be assessed very quickly. Behavioral synthesis tools are starting to appear that automate the process from about step 2 onwards. Figure 13.3 shows the overall design flow for RTL synthesis-based design.

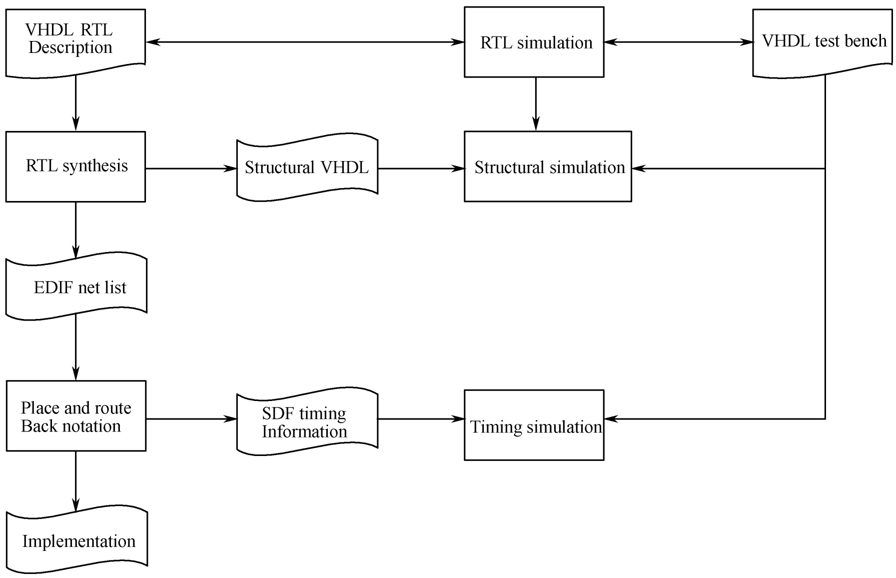


Figure 13.3 The overall design flow for RTL synthesis-based design

New Words

lifetime ['laiftaim] *n.* 寿命

lessen ['lesn] *v.* 减少,减轻

specialized ['speʃəlaizd] *adj.* 专门的,专用的

well-paid ['wel-'peid] *adj.* 收入高的

profession [prə'feʃn] *n.* 职业,专业

diagram ['daiəgræm] *n.* 图表

textual ['tekstjuəl] *adj.* 文本的

vendor ['vendə] *n.* 厂商

laborious [lə'bɔːriəs] *adj.* 费力的

cost-effective ['kɔst-i'fektiv] *adj.* 合算的

mass-produced ['mæs-prə'djuːst] *adj.* 大量生产的

inevitably [in'evitəbli] *adv.* 不可避免

daunting [dɔːntiŋ] *adj.* 使人畏缩的

uncommitted [ˌʌnkə'mitid] *adj.* 未确定用途的

methodology [meθə'dɔlədʒi] *n.* 方法学

specification [ˌspesifi'keiʃn] *n.* 技术要求,规格明细

partition [pɑː'tiʃn] *n.* 分割,划分

placement ['pleismənt] *n.* 布置,安排

Phrases & Expressions

in practical terms 实际上

in terms of 根据,在……方面

Technical Terms

synthesis ['sinθisis] *n.* 综合

simulation [ˌsimju'leiʃn] *n.* 模拟,仿真

schematic [ski'mætik] *n.* 原理图,示意图

capture ['kæptʃə] *v.* 记录,输入

simulator ['simjuleitə] *n.* 模拟器,仿真器

metalization [ˌmetə'laizeiʃn] *n.* 金属化

sequential [siˈkwenʃəl] *adj*. 时序的

asynchronous [eiˈsiŋkrənəs] *adj*. 异步的

synchronous [ˈsiŋkrənəs] *adj*. 同步的

trigger [ˈtrigə] *vt*. 触发

routing [ˈruːtiŋ] *n*. 布线

discrete [disˈkriːt] *adj*. 离散的

full-custom 全定制的

semi-custom 半定制的

mixed-signal 混合信号

embedded system 嵌入式系统

building block 构件,模块

metal interconnect 金属互联

design flow 设计流程

present state 现态

state machine 状态机

next state 次态

Boolean variable 布尔变量

behavioral synthesis 行为综合

CD *abbr*. Compact Disc 光盘

LP *abbr*. Long Playing (Record) 慢转密纹唱片

CAD *abbr*. Computer Aided Design 计算机辅助设计

EDA *abbr*. Electronic Design Automation 电子设计自动化

HDL *abbr*. Hardware Description Language 硬件描述语言

ASIC [ˈeisik] *abbr*. Application Specific Integrated Circuit 专用集成电路

PLD *abbr*. Programmable Logic Device 可编程逻辑器件

PLA *abbr*. Programmable Logic Array 可编程逻辑阵列

FPGA *abbr*. Field Programmable Gate Array 现场可编程门阵列

RTL *abbr*. Register Transfer Level 寄存器传送级

Notes

1. 此句可译为:电子设计自动化工具可以执行“综合”和“模拟”这两类任务。综合是将设计指标转化为实际的设计实现,而模拟是对设计指标或详细的实现方案进行“演习”以确保其运行正确。

2. 布尔(George Boole,1815－1864)是英国数学家和哲学家,“布尔代数”(Boolean algebra)的发明人。
3. 此句可译为:为逻辑门选择合适的布局——集成电路中包含哪些逻辑门以及各个集成电路在印制电路板上的位置。

Lesson 14　FPGAs

What are FPGAs?

Field programmable gate arrays (FPGAs) are digital integrated circuits (ICs) that contain configurable (programmable) blocks of logic along with configurable interconnects between these blocks. Design engineers can configure (program) such devices to perform a tremendous variety of tasks.

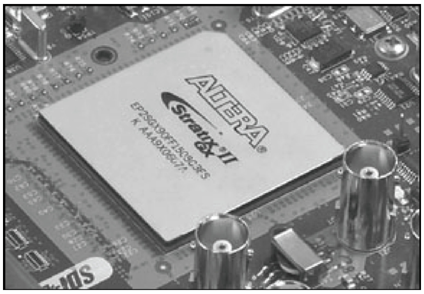


Figure 14.1　The Altera Stratix Ⅱ GX FPGA

Depending on the way in which they are implemented, some FPGAs may only be programmed a single time, while others may be reprogrammed over and over again. Not surprisingly, a device that can be programmed only one time is referred to as one-time programmable (OTP).

The “field programmable” portion of the FPGA's name refers to the fact that its programming takes place “in the field” (as opposed to devices whose internal functionality is hardwired by the manufacturer). This may mean that FPGAs are configured in the laboratory, or it may refer to modifying the function of a device resident in an electronic system that has already been deployed in the outside world. If a device is capable of being programmed while remaining resident in a higher-level system, it is referred to as being in-system programmable (ISP).

Why are FPGAs of interest?

There are many different types of digital ICs, including "jelly-bean logic" (small components containing a few simple, fixed logical functions), memory devices, and microprocessors (μPs). Of particular interest to us here, however, are programmable logic devices (PLDs), application-specific integrated circuits (ASICs), application-specific standard parts (ASSPs), and-of course-FPGAs.

For the purposes of this portion of our discussion, we shall consider the term PLD to encompass both simple programmable logic devices (SPLDs) and complex programmable logic devices (CPLDs) [1].

Various aspects of PLDs, ASICs, and ASSPs will be introduced in more detail in chapters 2 and 3. For the nonce, we need only be aware that PLDs are devices whose internal architecture is predetermined by the manufacturer, but which are created in such a way that they can be configured (programmed) by engineers in the field to perform a variety of different functions. In comparison to an FPGA, however, these devices contain a relatively limited number of logic gates, and the functions they can be used to implement are much smaller and simpler.

At the other end of the spectrum are ASICs and ASSPs, which can contain hundreds of millions of logic gates and can be used to create incredibly large and complex functions. ASICs and ASSPs are based on the same design processes and manufacturing technologies. Both are custom-designed to address a specific application, the only difference being that an ASIC is designed and built to order for use by a specific company, while an ASSP is marketed to multiple customers. (When we use the term ASIC henceforth, it may be assumed that we are also referring to ASSPs unless otherwise noted or where such interpretation is inconsistent with the context.)

Although ASICs offer the ultimate in size (number of transistors), complexity, and performance; designing and building one is an extremely time-consuming and expensive process, with the added disadvantage that the final design is "frozen in silicon" and cannot be modified without creating a new version of the device.

Thus, FPGAs occupy a middle ground between PLDs and ASICs because their functionality can be customized in the field like PLDs, but they can contain millions of logic gates and be used to implement extremely large and complex functions that previously could be realized only using ASICs.

The cost of an FPGA design is much lower than that of an ASIC (although the ensuing ASIC components are much cheaper in large production runs). At the same

time, implementing design changes is much easier in FPGAs, and the time-to-market for such designs is much faster. Thus, FPGAs make a lot of small, innovative design companies viable because-in addition to their use by large system design houses-FPGAs facilitate "Fred-in-the-shed"-type operations. This means they allow individual engineers or small groups of engineers to realize their hardware and software concepts on an FPGA-based test platform without having to incur the enormous nonrecurring engineering (NRE) costs or purchase the expensive toolsets associated with ASIC designs. Hence, there were estimated to be only 1,500 to 4,000 ASIC design starts and 5,000 ASSP design starts in 2003 (these numbers are falling dramatically year by year), as opposed to an educated "guesstimate" of around 450,000 FPGA design starts in the same year.

What can FPGAs be used for?

When they first arrived on the scene in the mid-1980s, FPGAs were largely used to implement glue logic, medium complexity state machines, and relatively limited data processing tasks. During the early 1990s, as the size and sophistication of FPGAs started to increase, their big markets at that time were in the telecommunications and networking arenas, both of which involved processing large blocks of data and pushing that data around. Later, toward the end of the 1990s, the use of FPGAs in consumer, automotive, and industrial applications underwent a humongous growth spurt.

FPGAs are often used to prototype ASIC designs or to provide a hardware platform on which to verify the physical implementation of new algorithms. However, their low development cost and short time-to-market mean that they are increasingly finding their way into final products (some of the major FPGA vendors actually have devices that they specifically market as competing directly against ASICs).

By the early-2000s, high-performance FPGAs containing millions of gates had become available. Some of these devices feature embedded microprocessor cores, high-speed input/output (I/O) interfaces, and the like. The end result is that today's FPGAs can be used to implement just about anything, including communications devices and software-defined radios (SDR); radar, image, and other digital signal processing (DSP) applications; all the way up to system-on-chip (SoC) components that contain both hardware and software elements.

To be just a tad more specific, FPGAs are currently eating into four major market segments: ASIC and custom silicon, DSP, embedded microcontroller applications, and physical layer communication chips. Furthermore, FPGAs have created a new market in

their own right: reconfigurable computing (RC).

- **ASIC and custom silicon**: As was discussed in the previous section, today's FPGAs are increasingly being used to implement a variety of designs that could previously have been realized using only ASICs and custom silicon.
- **Digital signal processing**: High—speed DSP has traditionally been implemented using specially tailored microprocessors called digital signal processors (DSPs). However, today's FPGAs can contain embedded multipliers, dedicated arithmetic routing, and large amounts of on-chip RAM, all of which facilitate DSP operations. When these features are coupled with the massive parallelism provided by FPGAs, the result is to outperform the fastest DSP chips by a factor of 500 or more.
- **Embedded microcontrollers**: Small control functions have traditionally been handled by special-purpose embedded processors called microcontrollers. These low-cost devices contain on-chip program and instruction memories, timers, and I/O peripherals wrapped around a processor core. FPGA prices are falling, however, and even the smallest devices now have more than enough capability to implement a soft processor core combined with a selection of custom I/O functions. The end result is that FPGAs are becoming increasingly attractive for embedded control applications.
- **Physical layer communications**: FPGAs have long been used to implement the glue logic that interfaces between physical layer communication chips and high-level networking protocol layers. The fact that today's high-end FPGAs can contain multiple high-speed transceivers means that communications and networking functions can be consolidated into a single device.
- **Reconfigurable computing**: This refers to exploiting the inherent parallelism and reconfigurability provided by FPGAs to "hardware accelerate" software algorithms. Various companies are currently building huge FPGA-based reconfigurable computing engines for tasks ranging from hardware simulation to cryptography analysis to discovering new drugs.

New Words

encompass [inˈkʌmpəs] *v.* 包含
spectrum [ˈspektrəm] *n.* 光谱,频谱,范围
address [əˈdres] *vt.* 从事,忙于

market [ˈmɑːkit] *n.* 市场，销路，行情
henceforth [hensˈfɔːθ] *adv.* 自此以后，今后
interpretation [inˌtəːpriˈteiʃn] *n.* 解释，阐明
inconsistent [ˌinkənˈsistənt] *adj.* 不一致的，矛盾的
context [ˈkɔntekst] *n.* 上下文，背景，环境
ultimate [ˈʌltimit] *adj.* 最终的，根本的
disadvantage [ˌdisədˈvɑːntidʒ] *n.* 缺点，劣势
ensue [inˈsjuː] *vi.* 跟着发生
viable [ˈvaiəbl] *adj.* 可行的
innovative [ˈinəuveitiv] *adj.* 创新的
shed [ʃed] *n.* 棚，小屋
incur [inˈkəː] *v.* 招致
educated [ˈedjuːkeitid] *adj.* 受过教育的，有教养的，有根据的
guesstimate [ˈgestimit] *n.* 估计，猜测
nonrecurring [ˈnɔnriˈkəːriŋ] *adj.* 一次性的，不重现的
arena [əˈriːnə] *n.* 竞技场，舞台
humongous [hjuːˈmʌŋgəs] *adj.* 极大的
spurt [spəːt] *n.* 喷射，迸发，冲刺
tailor [ˈteilə] *vt.* 剪裁，修改，调整
dedicate [ˈdedikeit] *vt.* 专用，致力于
facilitate [fəˈsiliteit] *vt.* 使容易，使便利
couple [ˈkʌpl] *vt.* 连接，结合
wrapp [ræp] *vt.* 包裹，覆盖，缠绕
consolidated [kənˈsɔlideitid] *adj.* 加固的，整理过的，统一的

Phrases & Expressions

be referred to as… 被称作……
in the field 在现场
as opposed to … 与……相反
for the nonce 目前，暂且
in one's own right 依靠自身的本领或素质

Technical Terms

parallelism [ˈpærəlelizəm] *n.* 并行度
prototype [ˈprəutətaip] *n.* 原型，样机
cryptography [kripˈtɔgrəfi] *n.* 密码系统，密码术
state machine 状态机
data processing 数据处理
glue logic 胶连逻辑
OTP *abbr.* One-Time Programmable 一次可编程
ISP *abbr.* In-System Programmable 在系统可编程
μP *abbr.* microprocessor 微处理器
PLD *abbr.* Programmable Logic Device 可编程逻辑器件
ASIC *abbr.* Application-Specific Integrated Circuit 专用集成电路
ASSP *abbr.* Application-Specific Standard Parts 专用标准器件
SPLD *abbr.* Simple Programmable Logic Devices 简单可编程逻辑器件
CPLD *abbr.* Complex Programmable Logic Devices 复杂可编程逻辑器件
NRE *abbr.* Nonrecurring Engineering 一次性工程
SDR *abbr.* Software-Defined Radios 软件无线电
DSP *abbr.* Digital Signal Processor 数字信号处理器
SoC *abbr.* System-on-Chip 片上系统
RC *abbr.* Reconfigurable Computing 可重配计算

Notes

1. CPLD 是一种复杂度在 PAL 和 FPGA 之间的可编程逻辑器件，其组成模块是“宏单元”(macro cell)。

Lesson 15　VHDL

What is VHDL?

VHDL is a programming language that has been designed and optimized for

describing the behavior of digital systems. VHDL has many features appropriate for describing the behavior of electronic components ranging from simple logic gates to complete microprocessors and custom chips. Features of VHDL allow electrical aspects of circuit behavior (such as rise and fall times of signals, delays through gates, and functional operation) to be precisely described. The resulting VHDL simulation models can then be used as building blocks in larger circuits (using schematics, block diagrams or system-level VHDL descriptions) for the purpose of simulation.

VHDL is also a general-purpose programming language: just as high-level programming languages allow complex design concepts to be expressed as computer programs, VHDL allows the behavior of complex electronic circuits to be captured into a design system for automatic circuit synthesis or for system simulation. Like Pascal, C and C++, VHDL includes features useful for structured design techniques, and offers a rich set of control and data representation features. Unlike these other programming languages, VHDL provides features allowing concurrent events to be described. This is important because the hardware described using VHDL is inherently concurrent in its operation.

One of the most important applications of VHDL is to capture the performance specification for a circuit, in the form of what is commonly referred to as a test bench. Test benches are VHDL descriptions of circuit stimuli and corresponding expected outputs that verify the behavior of a circuit over time. Test benches should be an integral part of any VHDL project and should be created in tandem with other descriptions of the circuit.

A standard language

One of the most compelling reasons for you to become experienced with and knowledgeable in VHDL is its adoption as a standard in the electronic design community. Using a standard language such as VHDL virtually guarantees that you will not have to throw away and recapture design concepts simply because the design entry method you have chosen is not supported in a newer generation of design tools. Using a standard language also means that you are more likely to be able to take advantage of the most up-to-date design tools and that you will have access to a knowledge base of thousands of other engineers, many of whom are solving problems similar to your own.

A brief history of VHDL

VHDL, which stands for VHSIC (Very High Speed Integrated Circuit) Hardware

Description Language, was developed in the early 1980s as a spin-off of a high-speed integrated circuit research project funded by the U. S. Department of Defense. During the VHSIC program, researchers were confronted with the daunting task of describing circuits of enormous scale (for their time) and of managing very large circuit design problems that involved multiple teams of engineers. With only gate-level design tools available, it soon became clear that better, more structured design methods and tools would be needed.

To meet this challenge, a team of engineers from three companies (IBM, Texas Instruments and Intermetrics) were contracted by the Department of Defense to complete the specification and implementation of a new, language-based design description method. The first publicly available version of VHDL, version 7. 2, was released in 1985. In 1986, the Institute of Electrical and Electronics Engineers, Inc. (IEEE)[1] was presented with a proposal to standardize the language, which it did in 1987 after substantial enhancements and modifications were made by a team of commercial, government and academic representatives. The resulting standard, IEEE 1076-1987, is the basis for virtually every simulation and synthesis product sold today. An enhanced and updated version of the language, IEEE 1076-1993, was released in 1994, and VHDL tool vendors have been responding by adding these new language features to their products.

Although IEEE Standard 1076 defines the complete VHDL language, there are aspects of the language that make it difficult to write completely portable design descriptions (descriptions that can be simulated identically using different vendors' tools). The problem stems from the fact that VHDL supports many abstract data types, but it does not address the simple problem of characterizing different signal strengths or commonly used simulation conditions such as unknowns and high-impedance.

Soon after IEEE 1076-1987 was adopted, simulator companies began enhancing VHDL with new, non-standard types to allow their customers to accurately simulate complex electronic circuits. This caused problems because design descriptions entered into one simulator were often incompatible with other simulation environments. VHDL was quickly becoming a non-standard.

To get around the problem of non-standard data types, another standard was developed by an IEEE committee. This standard, numbered 1164, defines a standard package (a VHDL feature that allows commonly used declarations to be collected into an external library) containing definitions for a standard nine-valued data type [2]. This standard data type is called std_logic, and the IEEE 1164 package is often referred to as the Standard Logic package.

The IEEE 1076-1987 and IEEE 1164 standards together form the complete VHDL

standard in widest use today. (IEEE 1076-1993 is slowly working its way into the VHDL mainstream, but it does not add significant new features for synthesis users.)

Standard 1076.3 (often called the Numeric Standard or Synthesis Standard) defines standard packages and interpretations for VHDL data types as they relate to actual hardware. This standard, which was released at the end of 1995, is intended to replace the many custom (non-standard) packages that vendors of synthesis tools have created and distributed with their products.

IEEE Standard 1076.3 does for synthesis users what IEEE 1164 did for simulation users: increase the power of Standard 1076, while at the same time ensuring compatibility between different vendors' tools [3]. The 1076.3 standard includes, among other things:

• A documented hardware interpretation of values belonging to the bit and Boolean types defined by IEEE Standard 1076, as well as interpretations of the std_ulogic type defined by IEEE Standard 1164.

• A function that provides "don't care" or "wild card" testing of values based on the std_ulogic type. This is of particular use for synthesis, since it is often helpful to express logic in terms of "don't care" values.

• Definitions for standard signed and unsigned arithmetic data types, along with arithmetic, shift, and type conversion operations for those types.

The annotation of timing information to a simulation model is an important aspect of accurate digital simulation. The VHDL 1076 standard describes a variety of language features that can be used for timing annotation. However, it does not describe a standard method for expressing timing data outside of the timing model itself.

The ability to separate the behavioral description of a simulation model from the timing specifications is important for many reasons. One of the major strengths of Verilog HDL (VHDL's closest rival) is the fact that Verilog HDL includes a feature specifically intended for timing annotation. This feature, the Standard Delay Format, or SDF, allows timing data to be expressed in a tabular form and included into the Verilog timing model at the time of simulation.

The IEEE 1076.4 standard, published by the IEEE in late 1995, adds this capability to VHDL as a standard package. A primary impetus behind this standard effort was to make it easier for ASIC vendors and others to generate timing models applicable to both VHDL and Verilog HDL. For this reason, the underlying data formats of IEEE 1076.4 and Verilog's SDF are quite similar.

When should you use VHDL?

Why choose to use VHDL for your design efforts? There are many likely reasons. If you ask most VHDL tool vendors this question, the first answer you will get is, "It will improve your productivity." But just what does this mean? Can you really expect to get your projects done faster using VHDL than by using your existing design methods?

The answer is yes, but probably not the first time you use it, and only if you apply VHDL in a structured manner. VHDL (like a structured software design language) is most beneficial when you use a structured, top-down approach to design. Real increases in productivity will come later, when you have climbed higher on the VHDL learning curve and have accumulated a library of reusable VHDL components.

Productivity increases will also occur when you begin to use VHDL to enhance communication between team members and when you take advantage of the more powerful tools for simulation and design verification that are available [4]. In addition, VHDL allows you to design at a more abstract level. Instead of focusing on a gate-level implementation, you can address the behavioral function of the design.

How will VHDL increase your productivity? By making it easy to build and use libraries of commonly-used VHDL modules. VHDL makes design reuse feel natural. As you discover the benefits of reusable code, you will soon find yourself thinking of ways to write your VHDL statements in ways that make them general purpose [5]. Writing portable code will become an automatic reflex.

Another important reason to use VHDL is the rapid pace of development in electronic design automation (EDA) tools and in target technologies. Using a standard language such as VHDL can greatly improve your chances of moving into more advanced tools (for example, from a basic low-cost simulator to a more advanced one) without having to re-enter your circuit descriptions. Your ability to retarget circuits to new types of device targets (for example, ASICs, FPGAs, and complex PLDs) will also be improved by using a standard design entry method.

New Words

appropriate [əˈprəupriət] *adj.* 适当的

concurrent [kənˈkʌrənt] *adj.* 并发的

integral [ˈintigrəl] *adj.* 整体的 *n.* 积分

tandem [ˈtændəm] *adj.* 级联的，串联的

virtually [ˈvəːtjuəli] *adv.* 事实上,实质上
compelling [kəmˈpeliŋ] *adj.* 强制的
proposal [prəˈpəuzəl] *n.* 提议,建议
enhancement [inˈhɑːnsmənt] *n.* 增强
modification [ˌmɔdifiˈkeiʃən] *n.* 更改,修正
unknown [ˈʌnˈnəun] *n.* 未知数,未知状态
declaration [ˌdekləˈreiʃn] *n.* 声明
mainstream [ˈmeinstriːm] *n.* 主流
annotation [ˌænəuˈteiʃn] *n.* 标注,注解
variety [vəˈraiəti] *n.* 种类
tabular [ˈtæbjulə] *adj.* 表格式的
productivity [ˌprɔdʌkˈtivəti] *n.* 生产力;生产率
beneficial [ˌbeniˈfiʃəl] *adj.* 有益的,受益的
up-to-date *adj.* 最近的,最新的
reusable [riːˈjuːzəbl] *adj.* 可重用的
statement [ˈsteitmənt] *n.* 语句
reflex [ˈriːfleks] *n.* 反射,本能反应

Phrases & Expressions

in tandem with 一前一后
have access to 可以到达,可以使用,可以理解
get around 绕过,避开
work one's way into 兢兢业业地达成

Technical Terms

stimuli [ˈstimjulai] *n.* 激励源
impedance [imˈpiːdəns] *n.* 阻抗
package [ˈpækidʒ] *n.* 芯片封装,程序包,成套设备
timing [ˈtaimiŋ] *n.* 时序
don't care *adj.* 无关的
wild card *n.* 通配符
test bench 测试台
high impedance 高阻

type conversion 类型转换
top-down approach 自顶向下的方法
learning curve 学习曲线
IEEE *abbr*. Institute of Electrical and Electronics Engineers 电气与电子工程师学会

Notes

1. 电气与电子工程师学会(IEEE)是一个国际性学术组织。1963 年,美国电气工程师学会和无线电工程师学会合并组成 IEEE。IEEE 总部设在纽约。
2. 此句可译为:该标准(标准号为 1164)定义了一个包含一个标准 9 值数据类型定义的标准包(标准包是 VHDL 的一个特点,即允许将共用数据类型声明集中在一个外部库函数中)。
3. 此句可译为:IEEE1076.3 标准为综合用户做的和 IEEE1169 标准为仿真用户做的是一致的:在确保兼容不同生产商工具的同时,提高 1076 标准的性能。
4. 此句可译为:当你为增进项目小组成员间的交流而开始使用 VHDL 时,当你利用更强大的工具进行仿真和设计验证时,工作效率就会提高。
5. 此句可译为:当你发现了可重用代码的好处之后,你很快会发觉自己正在考虑采用 VHDL 使自己编写的语句通用化。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) ______(随着 20 世纪 80 年代超大规模集成技术的到来) engineers began to realize the advantages of designing an IC that was customized or tailored to a particular system or application rather than using ______(标准集成电路) alone. Microelectronic system design then becomes a matter ______ defining the functions that you can ______(实现) using standard ICs and then implementing the remaining ______(逻辑功能) (sometimes called glue logic) with one or more ______(定制集成电路). As VLSI became possible, you could build a system from ______(数量更少的部件) by combining many standard ICs ______ a few custom ICs. Building a microelectronic system with fewer ICs allows you to ______(降低成本和提高可靠性).

Of course, there are many situations ______ which it is not appropriate ______ use a custom IC for each and every part of an microelectronic system. ______(如果需要大容量的存储器), for example, it is still best to use standard memory ICs, either dynamic random-access memory (DRAM or dRAM), or static RAM (SRAM or sRAM), ______

conjunction with custom ICs.

One of the first conferences to be devoted ______ this rapidly emerging segment of the IC industry was the IEEE Custom Integrated Circuits Conference (CICC), and the proceedings of this annual conference form a useful reference ______ the development of custom ICs. As different types of custom ICs began to evolve for different types of applications, these new ICs gave rise to a new term: application-specific IC, or ASIC. Now we have the IEEE International ASIC Conference, which tracks advances in ASICs separately from other types of custom ICs.

(2) PLD (Programmable Logic Device) is an umbrella term for ______(一类芯片) that are programmable at the customer's site ______ this case, the customer is the circuit developer, not the end user. There are three physical structures.

The first is the permanent fuse type that blows apart lines or fuses them together ______ electrically melting an aluminum trace or insulator. This was the first type of PLD, ______ "programmable array logic" or PAL.

The second is reprogrammable and uses EEPROM or flash memory. It causes a transistor ______ open or close ______ (依据) the contents of its associated memory cell.

The third type is RAM based, ______ makes it dynamic and volatile. Its contents are loaded ______(每次) it starts up. This control memory is not a self-contained memory-bank-like one that is used to input data. It is a separate masking layer in the chip, and each programmable point is connected ______ its counterpart bit in this layer.

CPLDs (Complex PLDs) and FPGAs (Field Programmable Gate Arrays) are ______(最常见的可编程逻辑器件芯片). CPLDs are mostly EEPROM or flash based. FPGAs use all three methods.

______(和门阵列不同), which require the final masking fabrication process, PLDs are easily programmable ______(在现场). PLDs are always used for logical functions, but ______(可编程存储芯片) such as PROMs and EPROMs might also be considered PLDs if they contain program code ______ just data.

2. Translate the following passages into Chinese or English.

1) Some digital logic ICs and their analog counterparts (analog/digital converters, for example) are standard parts, or standard ICs. You can select standard ICs from catalogs and data books and buy them from distributors. Systems manufacturers and designers can use the same standard part in a variety of different microelectronic systems (systems that use microelectronics or ICs).

2) A modern submicron CMOS process is now just as complicated as a submicron

bipolar or BiCMOS (a combination of bipolar and CMOS) process. However, CMOS ICs have established a dominant position, are manufactured in much greater volume than any other technology, and therefore, because of the economy of scale, the cost of CMOS ICs is less than a bipolar or BiCMOS IC for the same function. Bipolar and BiCMOS ICs are still used for special needs. For example, bipolar technology is generally capable of handling higher voltages than CMOS. This makes bipolar and BiCMOS ICs useful in power electronics, cars, telephone circuits, and so on.

3) An ASIC (Application Specific Integrated Circuit) is a chip that is custom designed for a specific application rather than a general-purpose chip such as a microprocessor. The use of ASICs improves performance over general-purpose CPUs, because ASICs are "hardwired" to do a specific job and do not incur the overhead of fetching and interpreting stored instructions. An ASIC chip performs an electronic operation as fast as it is possible to do so, providing, of course, that the circuit design is efficiently constructed.

4) A soft core is a block of logic for a particular function that is designed to be implemented in a programmable logic chip (PLD) or on a programmable section of a microcontroller chip or system on a chip (Sock). Depending on the vendor, the logic can come as a schematic, netlist or HDL code. Soft cores may be implemented with hard cores on a chip.

5) A hard core is a block of logic for a particular function on a chip that is designed at the circuit level. Microprocessors are typically hard cores, which ties them to a particular foundry's semiconductor process. Microcontrollers or systems on a chip (SoCs) may be entirely hard cores or made up of hard (hardwired) cores and soft (programmable) cores.

6)“电子设计自动化”就是利用计算机对芯片上的电子线路进行功能设计和仿真。

7)“硬件描述语言”是一种描述电子线路功能的语言,可编制文档、仿真或逻辑综合。

8)逻辑综合是将电子线路的高级描述转换为一张逻辑门及其互连列表(即“网表”)的过程。逻辑综合程序都能理解 Verilog 和 VHDL 的某个子集。

9)“复杂可编程逻辑器件”是一种在逻辑模块之间具有可编程互连的可编程逻辑器件。多数复杂可编程逻辑器件是基于电可擦写可编程只读存储器和闪存的。

10)“现场可编程门阵列”是一种由高密度门电路组成的可编程逻辑器件。现场可编程门阵列包含多达几十万个门电路,而且具有各种不同的结构。有的现场可编程门阵列非常复杂——不仅包括可编程逻辑模块,还包括可编程逻辑模块之间的可编程互连和开关。现场可编程门阵列多数是可重复编程的(基于电可擦写可编程只读存储器和闪存)或者是动态的(基于随机存储器)。

Reading Materials

Passage 1 Evolution of Programmable Logic Devices

PROMs

The first type of user-programmable chip that could implement logic circuits was the Programmable Read-Only Memory (PROM), in which address lines can be used as logic circuit inputs and data lines as outputs. Logic functions, however, rarely require more than a few product terms, and a PROM contains a full decoder for its address inputs. PROMs are thus an inefficient architecture for realizing logic circuits, and so are rarely used in practice for that purpose.

FPLAs

The first device developed later specifically for implementing logic circuits was the Field-Programmable Logic Array (FPLA), or simply PLA for short. A PLA consists of two levels of logic gates: a programmable "wired" AND-plane followed by a programmable "wired" OR-plane. A PLA is structured so that any of its inputs (or their complements) can be AND'ed together in the AND-plane; each AND-plane output can thus correspond to any product term of the inputs. Similarly, each OR-plane output can be configured to produce the logical sum of any of the AND-plane outputs. With this structure, PLAs are well-suited for implementing logic functions in sum of products form. They are also quite versatile, since both the AND terms and OR terms can have many inputs (this feature is often referred to as wide AND and OR gates).

When PLAs were introduced in the early 1970s, by Philips, their main drawbacks were that they were expensive to manufacture and offered somewhat poor speed-performance. Both disadvantages were due to the two levels of configurable logic, because programmable logic planes were difficult to manufacture and introduced significant propagation delays.

PALs

To overcome these weaknesses, Programmable Array Logic (PAL) devices were developed. As Figure 1 illustrates, PALs feature only a single level of programmability,

consisting of a programmable "wired" AND-plane that feeds fixed OR-gates. To compensate for lack of generality incurred because the OR-plane is fixed, several variants of PALs are produced, with different numbers of inputs and outputs, and various sizes of OR-gates. PALs usually contain flip-flops connected to the OR-gate outputs so that sequential circuits can be realized.

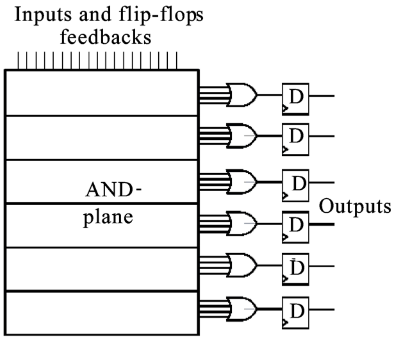


Figure 1 PALs structure

PAL devices are important because when introduced they had a profound effect on digital hardware design, and also they are the basis for some of the newer, more sophisticated architectures that will be described shortly. Variants of the basic PAL architecture are featured in several other products known by different acronyms. All small PLDs, including PLAs, PALs, and PAL-like devices are grouped into a single category called Simple PLDs (SPLDs), whose most important characteristics are low cost and very high pin-to-pin speed-performance.

CPLDs

As technology has advanced, it has become possible to produce devices with higher capacity than SPLDs. The difficulty with increasing capacity of a strict SPLD architecture is that the structure of the programmable logic-planes grow too quickly in size as the number of inputs is increased. The only feasible way to provide large capacity devices based on SPLD architectures is then to integrate multiple SPLDs onto a single chip and provide interconnect to programmably connect the SPLD blocks together. Many commercial FPD products exist on the market today with this basic structure, and are collectively referred to as Complex PLDs (CPLDs).

CPLDs were pioneered by Altera, first in their family of chips called Classic EPLDs, and then in three additional series, called MAX 5000, MAX 7000 and MAX 9000. Because of a rapidly growing market for large FPDs, other manufacturers

developed devices in the CPLD category and there are now many choices available. CPLDs provide logic capacity up to the equivalent of about 50 typical SPLD devices, but it is somewhat difficult to extend these architectures to higher densities. To build FPDs with very high logic capacity, a different approach is needed.

FPGAs

The highest capacity general purpose logic chips available today are the traditional gate arrays sometimes referred to as Mask-Programmable Gate Arrays (MPGAs). MPGAs consist of an array of prefabricated transistors that can be customized into the user's logic circuit by connecting the transistors with custom wires. Customization is performed during chip fabrication by specifying the metal interconnect, and this means that in order for a user to employ an MPGA a large setup cost is involved and manufacturing time is long. Although MPGAs are clearly not FPDs, they are mentioned here because they motivated the design of the user-programmable equivalent: Field-Programmable Gate Arrays (FPGAs).

Like MPGAs, FPGAs comprise an array of uncommitted circuit elements, called logic blocks, and interconnect resources, but FPGA configuration is performed through programming by the end user. An illustration of a typical FPGA architecture appears in Figure 2. As the only type of FPD that supports very high logic capacity, FPGAs have been responsible for a major shift in the way digital circuits are designed.

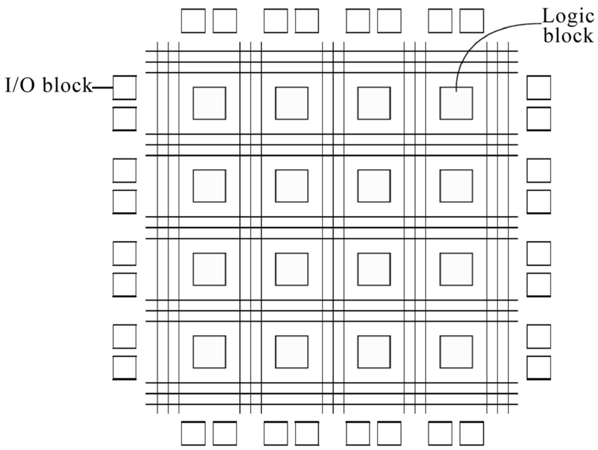


Figure 2 FPGAs structure

Figure 3 summarizes the categories of FPDs by listing the logic capacities available in each of the three categories. In the figure, "equivalent gates" refers loosely to

"number of 2-input NAND gates". The chart serves as a guide for selecting a specific device for a given application, depending on the logic capacity needed. However, each type of FPD is inherently better suited for some applications than for others. It should also be mentioned that there exist other special-purpose devices optimized for specific applications (e. g. state machines, analog gate arrays, large interconnection problems). However, since use of such devices is limited, they will not be described here.

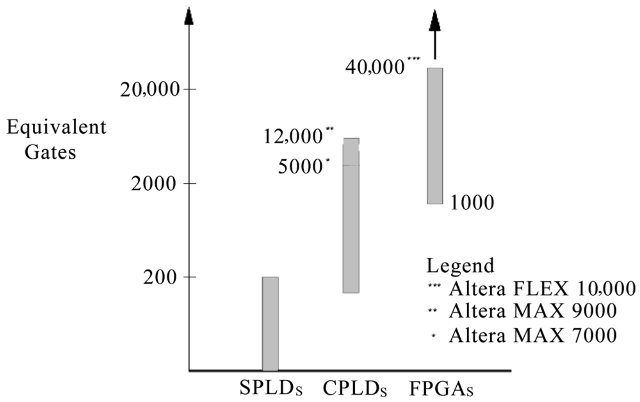


Figure 3 FPD categories by logic capacity

Questions:

1) Why are PROMs not efficient for realizing logic circuits?

2) What is the main difference between PLAs and PALs?

3) What does the term ***SPLD*** refer to?

4) What features does an typical FPGA device have ?

5) What does the term ***equivalent gates*** mean in this article?

Passage 2 Comparison of VHDL and Verilog

As the number of enhancements to various Hardware Description Languages (HDLs) has increased over the past year, so too has the complexity of determining which language is best for a particular design. Many designers and organizations are contemplating whether they should switch from one HDL to another. This paper compares the technical characteristics of two, general-purpose HDLs:

VHDL (IEEE-Std 1076): A general-purpose digital design language supported by

multiple verification and synthesis (implementation) tools.

Verilog (IEEE-Std 1364): A general-purpose digital design language supported by multiple verification and synthesis tools.

VHDL

VHDL is a strongly and richly typed language. Derived from the Ada programming language, its language requirements make it more verbose than Verilog. The additional verbosity is intended to make designs self-documenting. Also, the strong typing requires additional coding to explicitly convert from one data type to another (integer to bit-vector, for example). The creators of VHDL emphasized semantics that were unambiguous and designs that were easily portable from one tool to the next. Hence, race conditions, as an artifact of the language and tool implementation, are not a concern for VHDL users.

Several related standards have been developed to increase the utility of the language. Any VHDL design today depends on at least IEEE-Std 1164 (std_logic type), and many also depend on standard Numeric and Math packages as well. The development of related standards is due to another goal of VHDL's authors: namely, to produce a general language and allow development of reusable packages to cover functionality not built into the language.

VHDL does not define any simulation control or monitoring capabilities within the language. These capabilities are tool dependent. Due to this lack of language-defined simulation control commands and also because of VHDL's user-defined type capabilities, the VHDL community usually relies on interactive GUI environments for debugging design problems.

Verilog

Verilog is a weakly and limited typed language. Its heritage can be traced to the C programming language and an older HDL called Hilo. All data types in Verilog are predefined in the language. Verilog recognizes that all data types have a bit-level representation. The supported data representations (excluding strings) can be mixed freely in Verilog.

Simulation semantics in Verilog are more ambiguous than in VHDL. This ambiguity gives designers more flexibility in applying optimizations, but it can also (and often does) result in race conditions if careful coding guidelines are not followed. It is possible to have a design that generates different results on different vendors' tools or

even on different releases of the same vendor's tool.

Unlike the creators of VHDL, Verilog's authors thought that they provided designers everything they would need in the language. The more limited scope of the language combined with the lack of packaging capabilities makes it difficult, if not impossible, to develop reusable functionality not already included in the language.

Verilog defines a set of basic simulation control capabilities (system tasks) within the language. As a result of these predefined system tasks and a lack of complex data types, Verilog users often run batch or command-line simulations and debug design problems by viewing waveforms from a simulation results database.

Questions:

1) Where does VHDL stem from?

2) Can you tell some of the main goals of VHDL's creators?

3) Which language is Verilog derived from?

4) What's the meaning of ***ambiguity*** in this article? Why is it important?

5) What difference do you think the most considerable between VHDL and Verilog?

Passage 3 SoC

System-on-a-Chip (SoC) refers to integrating all components of a computer or other electronic system into a single integrated circuit (chip). It may contain digital, analog, mixed-signal, and often radio-frequency functions — all on one chip. A typical application is in the area of embedded systems.

If it is not feasible to construct a SoC for a particular application, an alternative is a system in package (SiP) comprising a number of chips in a single package. SoC is believed to be more cost effective since it increases the yield of the fabrication and because its packaging is simpler.

A typical SoC consists of the following blocks (see Figure 1):

- One or more microcontroller, microprocessor or DSP core(s).
- Memory blocks including a selection of ROM, RAM, EEPROM and Flash.
- Timing sources including oscillators and phase-locked loops (PLLs).
- Peripherals including counter-timers, real-time timers and power-on reset generators.
- External interfaces including industry standards such as USB, FireWire,

Ethernet, USART, SPI.

- Analog interfaces including ADCs and DACs.
- Voltage regulators and power management circuits.

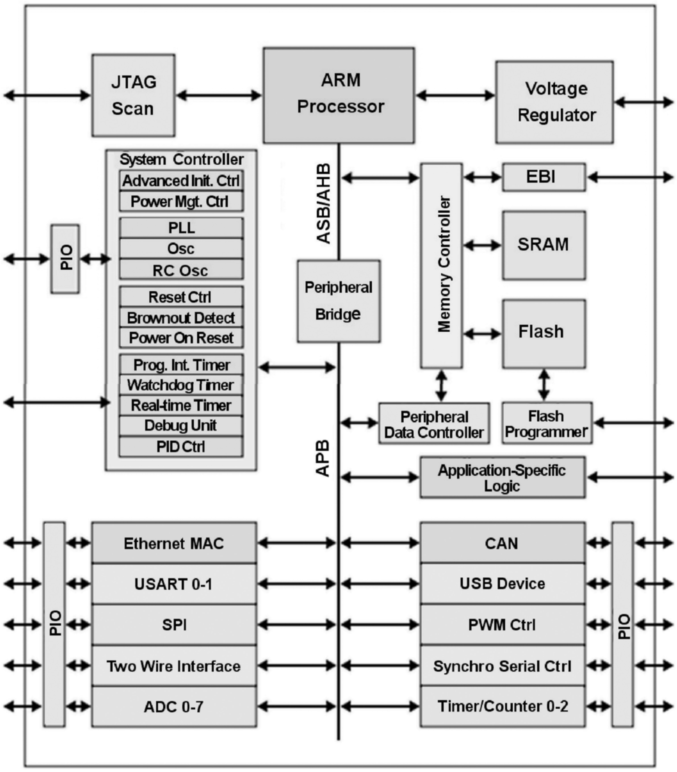


Figure 1 A microcontroller-based SoC

These blocks are connected by either a proprietary or industry-standard bus such as the AMBA bus from ARM. DMA controllers route data directly between external interfaces and memory, by-passing the processor core and thereby increasing the data throughput of the SoC.

A SoC consists of the hardware described above, and the software that controls the microcontroller, microprocessor or DSP cores, peripherals and interfaces. The design flow for an SoC aims to develop this hardware and software in parallel (see Figure 2).

Most SoCs are developed from pre-qualified hardware blocks for the hardware elements described above, together with the software drivers that control their operation. Of particular importance are the protocol stacks that drive industry-standard interfaces like USB. The hardware blocks are put together using CAD tools; the

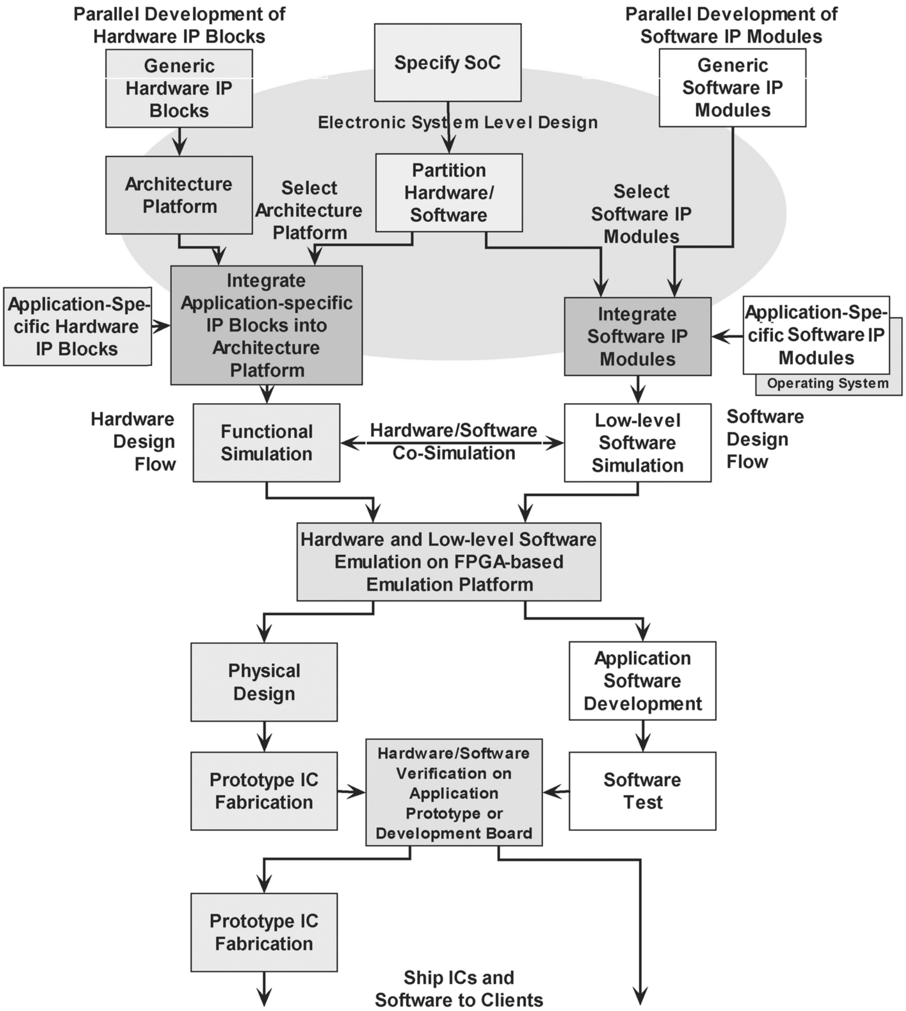


Figure 2 The SoC design flow

software modules are integrated using a software development environment.

A key step in the design flow is emulation: the hardware is mapped onto an emulation platform based on a field programmable gate array (FPGA) that mimics the behavior of the SoC, and the software modules are loaded into the memory of the emulation platform. Once programmed, the emulation platform enables the hardware and software of the SoC to be tested and debugged at close to its full operational speed. After emulation, the hardware of the SoC follows the place and route phase of the design of an integrated circuit before it is fabricated.

Chips are verified for logical correctness before being sent to foundry. The process is called ASIC verification. Verilog and VHDL are typical hardware description languages used for verification. With growing complexity of chips, hardware verification languages like SystemVerilog, SystemC and OpenVera are used. The bugs found in the verification stage are reported to the designer. Traditionally, 70% of time and energy in chip design life cycle are spent on verification.

SoCs can be fabricated by several technologies-full custom, standard cell and FPGA. SoC designs usually consume less power and have a lower cost and higher reliability than the multi-chip systems that they replace. And with fewer packages in the system, assembly costs are reduced as well. However, like most VLSI designs, the total cost is higher for one large chip than for the same functionality distributed over several smaller chips, because of lower yields and higher NRE costs.

Questions:

1) What does ***SoC*** and ***SIP*** stand for respectively?
2) What components does a typical SoC have?
3) What role does the FPGA device play in the SoC design flow?
4) Why is the phase of verification important in SoC design?
5) What does the term ***NRE*** mean in this article?

Unit 6

Digital Signal Processing

Lesson 16 Basic Concepts of DSP

We don't speak in a digital signal. A digital signal is a language of 1s and 0s that can be processed by mathematics. We speak in real-world, analog signals. Analog signals are real world signals that we experience everyday-sound, light, temperature, and pressure. A digital signal is a numerical representation of the analog signal. It may be easier and more cost effective to process these signals in the digital world. In the real world, we can convert these signals into digital signals through the analog-to-digital converter, process the signals, and if needed, bring the signals back out to the analog world through the digital-to-analog converter.

The essentials of analog-to-digital and digital-to-analog conversion

The first essential step in analog-to-digital (A/D) conversion is to sample an analog signal(See Figure 16.1). This step is performed by a sample and hold circuit, which samples at regular intervals called sampling intervals. The length of the sampling interval is the same as the sampling period, and the reciprocal of the sampling period is the sampling frequency fs. According to the Nyquist theorem, a signal with a maximum frequency of W Hz (called a band-limited signal) must be sampled at least 2 W samples per second to ensure accurate recording. When this minimum is not respected, distortion called aliasing occurs. Aliasing causes high frequency signals to appear as lower frequency signals. To be sure aliasing will not occur, sampling is always preceded by low pass filtering. The low pass filter, called the anti-aliasing filter, removes all frequencies above half the selected sampling rate.

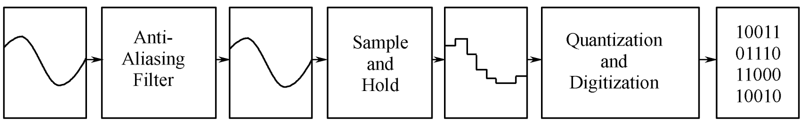


Figure 16.1 Analog-to-digital conversion

After a brief acquisition time, during which a sample is acquired, the sample and hold circuit holds the sample steady for the remainder of the sampling interval. This hold time is needed to allow time for an A/D converter to generate a digital code that best corresponds to the analog sample.

The A/D converter chooses a quantization level for each analog sample. An N-bit converter chooses among 2^N possible quantization levels. The larger the number of levels, the smaller the quantization errors, calculated as the difference between the quantized level and the true sample level. Most quantization errors are limited in size to half a quantization step Q. The quantization step size is calculated as $Q = R/2^N$, where R is the full scale range of the analog signal and N is the number of bits used by the converter. The strength of the signal compared to that of the quantization errors is measured by dynamic range [1] and signal-to-noise ratio.

A digital signal is represented by a set of vertical lines with circles at the top to mark the quantization levels selected for each sample. The bit rate for an A/D converter is the Nf_s, where fs is the sampling rate.

Finally, each digital sample is assigned a digital code, which completes the A/D process. The result is a digital bit stream. It is this collection of digital codes that is processed in digital signal processing.

To summarize, A/D comprises anti-aliasing, sampling, quantization and digitisation.

Once digital signal processing is complete, digital-to-analog (D/A) conversion must occur (See Figure 16.2). This process begins by converting each digital code into an analog voltage that is proportional in size to the number represented by the code. This voltage is held steady through zero order hold until the next code is available, one sampling interval later. This creates a staircase-like signal that contains frequencies above W Hz. These signals are removed with a smoothing low pass filter, the last step in D/A conversion.

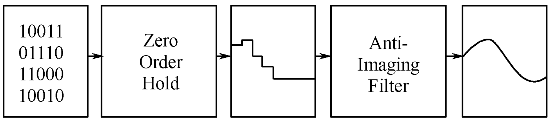


Figure 16.2 Digital-to-analog conversion

The images of each frequency f present in a sampled signal appear, through sampling, at the infinite number of frequencies $kf_s \pm f$Hz. When the sampling rate is lower than the required Nyquist rate, that is $f_s < 2W$, images of high frequency signals erroneously appear in the baseband [2] (or Nyquist range) due to aliasing. While this undersampling is normally avoided, it can be exploited. For example, signals whose frequencies are restricted to a narrow band of high frequencies can be sampled at a rate similar to twice the width of the band instead of twice the maximum frequency. All of the important signal characteristics can be deduced from the copy of the spectrum that appears in the baseband through sampling. Depending on the relationship between the

signal frequencies and the sampling rate, spectral inversion may cause the shape of the spectrum in the baseband to be inverted from the true spectrum of the signal.

Technologies for digital signal processing

If a universal microprocessor solution existed with which every design could be realized, the electronics industry wouldn't be a very competitive place[3]. However, typically in most electronic designs, more than one processor technology can be used to implement the required functions. The trick is, of course, to choose the one that best delivers the performance, size, power consumption, features, software and tools to get the job done fast — without breaking the budget. After almost two decades of development, digital signal processors continue to take the place of competitive processors. Digital signal processors are, after all, at the center of signal processing.

A digital signal processor (DSP) is a type of microprocessor—one that is incredibly fast and powerful. A DSP is unique because it processes data in real time. This real-time capability makes a DSP perfect for applications that cannot tolerate any delays. For example, did you ever talk on a cell phone where two people couldn't talk at once? You had to wait until the other person finished talking. If you both spoke simultaneously, the signal was cut—you didn't hear the other person. With today's digital cell phones, which use DSP, you can talk normally. The DSP processors inside cell phones process sounds so rapidly you hear them as quickly as you can speak - in real time. Here are just some of the advantages of designing with DSPs over other microprocessors:

- Single-cycle multiply-accumulate operations
- Real-time performance simulation and emulation[4]
- Flexibility
- Reliability
- Increased system performance
- Reduced system cost

However, there are some of the other alternatives available for digital signal processing. How they compare to DSPs?

The FPGA Alternative

Field-Programmable Gate Arrays have the capability of being reconfigurable within a system, which can be a big advantage in applications that need multiple trial versions within development, offering reasonably fast time to market. They also offer greater raw performance per specific operation because of the resulting dedicated logic circuit. However, FPGAs are significantly more expensive and typically have much higher

power dissipation than DSPs with similar functionality. As such, even when FPGAs are the chosen performance technology in designs such as wireless infrastructure, DSPs are typically used in conjunction with FPGAs to provide greater flexibility, better price/performance ratios, and lower system power.

The ASIC Alternative

Application-specific ICs can be tailored to perform specific functions extremely well, and can be made quite power efficient. However, since ASICS are not field-programmable, their functionality cannot be iteratively changed or updated while in product development. As such, every new version of the product requires a redesign and trips through the foundry[5], an expensive proposition, and an impediment to rapid time-to-market. Programmable DSPs, on the other hand, can be updated without changing the silicon, merely change the software program, greatly reducing development costs, and availing aftermarket feature enhancements with mere code downloads. Consequently, more often than not, when you see ASICs in real time signal processing applications, they are typically employed as bus interfaces, glue logic[6], and/or functional accelerators for a programmable DSP-based system.

The GPP Alternative

In contrast to ASICs that are optimised for specific functions, general-purpose microprocessors (GPPs) are best suited for performing a broad array of tasks. However, for applications in which the end product must process answers in real time, or must do so while powered by consumer batteries, GPPs comparatively poor real time performance and high power consumption all but rules them out [7]. More and more, these processors are being seen as the dinosaurs of the industry, too encumbered with PC compatibility and desktop features to adapt to the changing real time market place. As the world embraces tiny hand-held wireless-enabled products that require power dissipation measured in milliwatts-not the watts that these processors consume-DSPs are the programmable technology of choice. That trend is bound to continue as digital Internet appliances get smaller, faster and more portable.

New Words

numerical [nju:'merikl] *adj*. 数值的

reciprocal [ri'siprəkl] *n*. 倒数

precede [pri'si:d] *v*. 领先于

proportion [prə'pɔ:ʃn] *n*. 比例

smoothing [ˈsmuːðiŋ] *n.* 平滑
infinite [ˈinfinit] *adj.* 无限的
erroneous [iˈrəunjəs] *adj.* 错误的
deduce [diˈdjuːs] *vt.* 推导，演绎
incredibly [inˈkredəbli] *adv.* 难以置信地，惊人地
budget [ˈbʌdʒit] *n.* 预算
trick [trik] *n.* 窍门，诀窍
simultaneously [ˈsimlˈteiniəsli] *adv.* 同时地
alternative [ɔːlˈtəːnətiv] *n.* 选择
reconfigurable [ˈriːkənfiɡrəbl] *adj.* 可重新配置的
iterative [ˈitərətiv] *adj.* 重复的，迭代的
proposition [ˌprɔpəˈziʃn] *n.* 主张，建议
impediment [imˈpedimənt] *n.* 妨碍，阻碍
encumber [inˈkʌmbə] *v.* 阻碍
compatibility [kəmˌpætiˈbiliti] *n.* 兼容性
embrace [imˈbreis] *vt.* 拥抱，包含
appliance [əˈplaiəns] *n.* 用具，器具

Phrases & Expressions

in conjunction with 与……协力
more often than not 时常
be encumbered with 为……所累

Technical Terms

aliasing [ˈeiliəsiŋ] *n.* 混叠现象
digitisation [didʒitaiˈzeiʃn] *n.* 数字化
baseband [ˈbeisbænd] *n.* 基带
spectrum [ˈspektrəm] *n.* 频谱
undersampling [ˈʌndəˈsæmpliŋ] *n.* 欠采样
real time *n.* 实时
simulation [ˌsimjuˈleiʃn] *n.* 模拟
emulation [ˌemjuˈleiʃn] *n.* 仿真
field-programmable *adj.* 现场可编程的

foundry ['faundri] *n.* 半导体制造商
hand-held *adj.* 手持的,手持式的
sample and hold circuit 采样保持电路
sampling interval 采样间隔
Nyquist theorem 奈奎斯特定理
anti-aliasing filter 抗混叠滤波器
acquisition time 采集时间
quantization level 量化电平
full scale range 满量程范围
dynamic range 动态范围
signal-to-noise ratio 信噪比
zero order hold 零阶保持
spectral inversion 频谱反转
time to market 上市时间
wireless infrastructure 无线基础设施
price/performance ratio 性能价格比
bus interface 总线接口
glue logic 胶连逻辑
functional accelerator 性能加速器
end product 最终产品
power dissipation 功耗

Notes

1. 动态范围指的是信号最小值和最大值之间的范围,通常用信号最大值和信号最小值之比来表示。
2. 基带是指不加任何调制过程的数字信号传输。基带带宽全部用来传送数字脉冲信号,并可以通过时分复用的方式进行多路数字信号的传输。
3. 本句为虚拟语气。可译为:假如存在可用以实现任何设计的通用微处理器,电子行业就不会竞争得这么激烈了。
4. 模拟(simulation)和仿真(emulation)。模拟一般是指用软件的方法模拟微处理器的功能,而仿真是用硬件设备 — 仿真器(emulator)— 监测微处理器的实时运行。
5. foundry 是指为第三方制造芯片的半导体制造商,可以指那些出售其富余生产能力的大型芯片制造企业,也可以是指专为其他公司制造芯片的企业。
6. 胶连逻辑电路(glue logic)是一种用于将多个现成电路进行互连的定制数字电路。

胶连逻辑电路可以用 PLD 编程实现。

7. 本句可译为:然而,对于最终产品必须做出实时响应或者必须在电池供电下做出实时响应的应用而言,通用微处理器较差的实时性能和高功耗几乎将其排除在考虑之外。all but 此处作 rule them out 的状语,含义为“几乎,差一点”。rule out 的意思是“将……排除在外”。

Lesson 17 Digital Signal Processors

Digital signal processing tasks can be performed by all processors. Specialized digital signal processors (DSPs), however, perform these tasks most efficiently and most quickly. While traditional processors follow the Von Neumann architecture[1] model(Figure 17.1), which assumes a shared single memory to be used for both program instructions and data, DSPs use the Harvard or modified Harvard architecture[2], which includes multiple program and data memories, along with multiple buses to access them(Figure 17.2). This arrangement means that much less waiting is required when instructions or numbers are fetched from memory. In fact at least one of each can be fetched simultaneously. Such overlapping of tasks is called pipelining. In addition to multiple memories and buses, all DSPs have fast multipliers, accumulators, and shifters, and many have hardware support for circular buffers. Address generators can speed up accesses to memory locations referenced by registers.

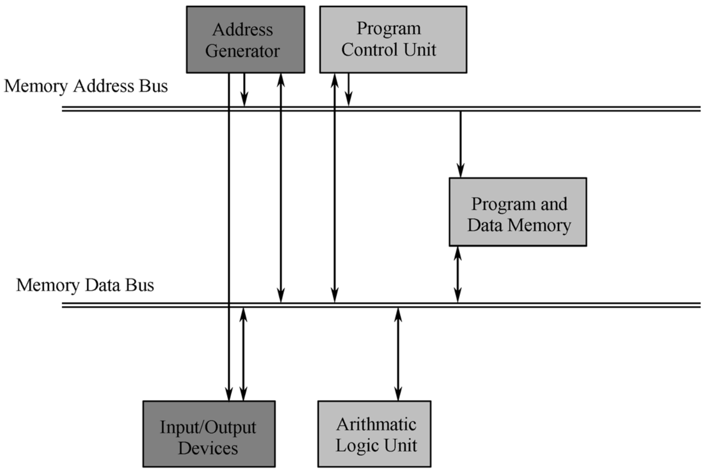


Figure 17.1 Von Neumann architecture

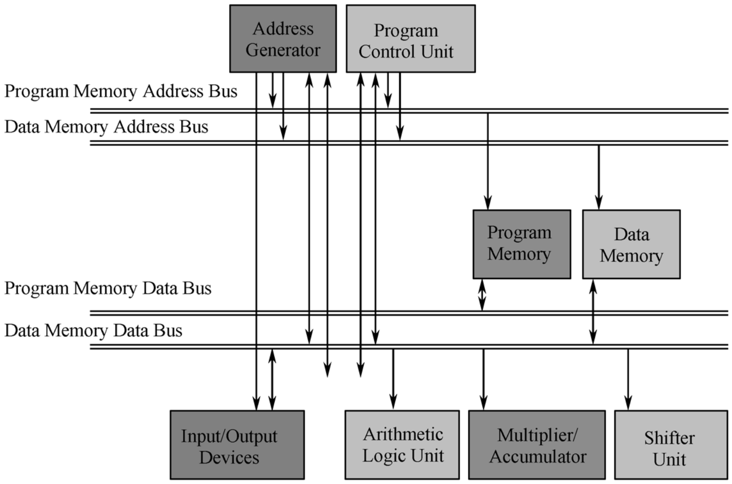


Figure 17.2 Harvard architecture

DSPs are available in two major classes: fixed point and floating point. The fixed point class represents real numbers in a fixed number of bits. The position of the binary point (similar to the decimal point) can be controlled by the programmer, and determines the range of numbers that can be represented. As the range increases, though, the available precision goes down, since fewer bits lie to the right of the binary point. In 16 bits, the formats 16.0, 15.1, 14.2, 13.3, 12.4, 11.5, 10.6, 9.7, 8.8, 7.9, 6.10, 5.11, 4.12, 3.13, 2.14, and 1.15 are possible. The dynamic range, calculated as 20log(Full Scale Range/Smallest Resolvable Difference), remains the same for all 16-bit formats, $20\log 2^{16} = 96.3$ dB.

Floating point DSPs represent real numbers using a mantissa and an exponent, similar to scientific notation: Many combine mantissa and exponent into a 32-bit number. The dynamic range for floating point devices is calculated from the largest and smallest multipliers 2^E, where E is the exponent. Thus, for a representation that uses 24 bits for the mantissa and 8 bits for the signed exponent, the dynamic range is $20\log(2^{127}/2^{-128}) = 1535.3$ dB. A large dynamic range means the system has great power to represent a wide range of input signals, from very small to very large.

Assembly language is the command language for DSPs. DSPs often have specialized instructions that make programming for common DSP tasks more convenient and more efficient. For example, most DSPs offer multi-function instructions that exploit their parallel architecture. Other constructs that are frequently offered are efficient looping schemes, since so many DSP operations involve a great deal of repetition.

Choosing a DSP for a particular application is not always easy. The first decision is on whether to choose a fixed point or a floating point device[3]. Generally, fixed point devices are cheaper and quicker, but floating point devices are more convenient to program and more suited to calculation-intensive algorithms. Second, the data width of the DSP determines how accurately it can represent numbers. Speed is another issue, not only how many cycles occur in each second, but also how many instructions execute in each cycle and how much work each of these instructions accomplishes. One way to assess the minimum requirements for the DSP is to estimate how many instructions must be executed for each received sample. When this number is multiplied by the sampling frequency, the minimum required number of instructions per second is obtained.

The specific hardware and software features offered by a particular DSP can make one choice better than another, as can the amount of on-chip memory available[4]. Sometimes DSPs are chosen because well-matched supporting hardware, particularly A/D and D/A converters, is obtainable. Frequently, the quality and convenience of the software tools, for both low level and high level programming languages, are also major factors, as is the availability of third party software. As always, cost is a factor. In fact, quite often, the DSP that is fastest and offers the most features, but also fits the budget, is the one selected.

DSPs can be purchased in three forms, as a core, as a processor, and as a board level product. In DSP, the term "core" refers to the section of the processor where the key tasks are carried out, including the data registers, multiplier, ALU, address generator, and program sequencer. A complete processor requires combining the core with memory and interfaces to the outside world. While the core and these peripheral sections are designed separately, they will be fabricated on the same piece of silicon, making the processor a single integrated circuit.

Suppose you build cellular telephones and want to include a DSP in the design. You will probably want to purchase the DSP as a processor, that is, an integrated circuit that contains the core, memory and other internal features. To incorporate this IC in your product, you have to design a printed circuit board where it will be soldered in next to your other electronics. This is the most common way that DSPs are used.

Now, suppose the company you work for manufactures its own integrated circuits. In this case, you might not want the entire processor, just the design of the core. After completing the appropriate licensing agreement, you can start making chips that are highly customized to your particular application. This gives you the flexibility of selecting how much memory is included, how the chip receives and transmits data, how

it is packaged, and so on. Custom devices of this type are an increasingly important segment of the DSP marketplace.

There are several dozen companies that will sell you DSPs already mounted on a printed circuit board. These have such features as extra memory, A/D and D/A converters, EPROM sockets, multiple processors on the same board, and so on. While some of these boards are intended to be used as stand alone computers, most are configured to be plugged into a host, such as a personal computer. Companies that make these types of boards are called Third Party Developers. The best way to find them is to ask the manufacturer of the DSP you want to use. Look at the DSP manufacturer's website; if you don't find a list there, send them an e-mail. They will be more than happy to tell you who are using their products and how to contact them.

Keep in mind that the distinction between DSPs and other microprocessors is not always a clear line. For instance, look at how Intel describes the MMX technology addition to its Pentium processor: "Intel engineers have added 57 powerful new instructions specifically designed to manipulate and process video, audio and graphical data efficiently. These instructions are oriented to the highly parallel, repetitive sequences often found in multimedia operations."

In the future, we will undoubtedly see more DSP-like functions merged into traditional microprocessors and microcontrollers. The Internet and other multimedia applications are a strong driving force for these changes. These applications are expanding so rapidly, in twenty years it is very possible that the Digital Signal Processor may be the "traditional" microprocessor.

New Words

specialized [ˈspeʃəlaizd] *adj*. 专门的,专用的

shared [ʃɛəd] *adj*. 共享的

circular [ˈsəːkjələ] *adj*. 循环的,环形的

reference [ˈrefrəns] *vt*. 定位 *n*. 访问,索引

fixed point *adj*. 定点的

floating point *adj*. 浮点的

well-matched *adj*. 非常匹配的

solder [ˈsɔldə] *n*. 焊料 *v*. 焊接

customize [ˈkʌstəmaiz] *v*. 定制

merge [məːdʒ] *v*. 合并,融合

Phrases & Expressions

in addition to 除……之外
as always 照常

Technical Terms

multiplier [ˈmʌltiplaiə] *n.* 乘法器
accumulator [əˈkjuːmjuleitə] *n.* 累加器
shifter [ˈʃiftə] *n.* 移位器
mantissa [mænˈtisə] *n.* 尾数
exponent [ikˈspəunənt] *n.* 指数
Von Neumann architecture 冯·诺伊曼结构
Harvard architecture 哈佛结构
circular buffer 循环缓冲区
address generator 地址产生器
smallest resolvable difference 最小可分辨值
scientific notation 科学记数法
parallel architecture 并行结构
looping scheme 循环机制
calculation-intensive algorithm 运算密集型算法
licensing agreement 专利使用权转让协定
third party developer 第三方开发商

Notes

1. "冯·诺伊曼结构"取名自美国杰出的数学家 — 约翰·冯·诺伊曼(John Von Neumann,1903—1957)。他引导了 20 世纪初许多重大数学发现。他的主要成就包括:提出了存储程序计算机(stored program computer)的概念、对量子力学的数学公式化及在原子弹方面的工作。
2. "哈佛结构"取名自 20 世纪 40 年代 Howard Aiken (1900—1973)领导的在哈佛大学(Harvard University)做的研究工作。
3. 此句可译为:首先要决定的是选择定点器件还是浮点器件。
4. 此句可译为:和片内可用存储器大小能做出更佳选择的判断一样,特定数字信号处理器所提供的软硬件功能会使一个选择优于另外一个选择。

Lesson 18 Comparison of DSP and ASP

Signals may be processed using analog techniques (analog signal processing, or ASP), digital techniques (digital signal processing, or DSP), or a combination of analog and digital techniques (mixed signal processing, or MSP). In some cases, the choice of techniques is clear; in others, there is no clear cut choice, and second-order considerations may be used to make the final decision[1].

With respect to DSP, the factor that distinguishes it from traditional computer analysis of data is its speed and efficiency in performing sophisticated digital processing functions such as filtering, FFT analysis, and data compression in real time.

The term mixed signal processing implies that both analog and digital processing is done as part of the system. The system may be implemented in the form of a printed circuit board or a single integrated circuit chip. In the context of this broad definition, ADCs and DACs are considered to be mixed signal processors, since both analog and digital functions are implemented in each. Recent advances in Very Large Scale Integration (VLSI) processing technology allow complex digital processing as well as analog processing to be performed on the same chip. The very nature of DSP itself implies that these functions can be performed in real-time.

ASP vs. DSP

Today's engineer faces a challenge in selecting the proper mix of analog and digital techniques to solve the signal processing task at hand. It is impossible to process real-world analog signals using purely digital techniques, since all sensors (microphones, thermocouples, strain gages, piezoelectric crystals, disk drive head s, etc.) are analog sensors. Therefore, some sort of signal conditioning circuitry is required in order to prepare the sensor output for further signal processing, whether it be analog or digital. Signal conditioning circuits are, in reality, analog signal processors, performing such functions as multiplication (gain), isolation (instrumentation amplifiers and isolation amplifiers), detection in the presence of noise (high common-mode instrumentation amplifiers etc.), dynamic range compression (log amps, LOGDACs, and programmable gain amplifiers), and filtering (both passive and active)[2]. Several methods of accomplishing signal processing are shown in Figure 18.1. The top portion of the figure

shows the purely analog approach. The latter parts of the figure show the DSP approach. Note that once the decision has been made to use DSP techniques, the next decision must be where to place the ADC in the signal path.

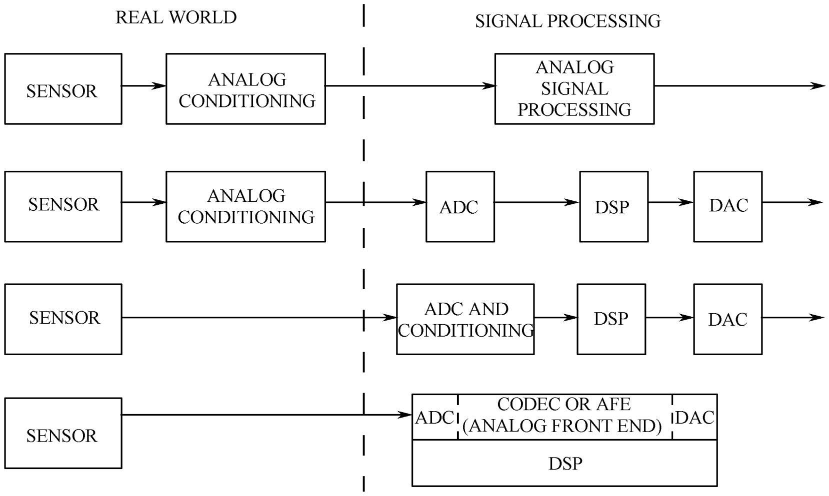


Figure 18.1 some methods of accomplishing signal processing

In general, as the ADC is moved closer to the actual sensor, more of the analog signal conditioning burden is now placed on the ADC. The added ADC complexity may take the form of increased sampling rate, wider dynamic range, higher resolution, input noise rejection, input filtering and programmable gain amplifiers (PGAs), on-chip voltage references, etc. , all of which add functionality and simplify the system[3]. With today's high-resolution/high sampling rate data converter technology, significant progress has been made in integrating more and more of the conditioning circuitry within the ADC/DAC itself. In the measurement area, for instance, 24-bit ADCs are available with built-in programmable gain amplifiers (PGAs) which allow full-scale bridge signals of 10mV to be digitized directly with no further conditioning. At voiceband and audio frequencies, complete coder-decoders (Codecs or Analog Front Ends) are available which have sufficient on-chip analog circuitry to minimize the requirements for external conditioning components. At video speeds, analog front ends are also available for such applications as CCD image processing and others.

A PRACTICAL EXAMPLE

As a practical example of the power of DSP, consider the comparison between an analog and a digital lowpass filter, each with a cutoff frequency of 1kHz. The digital

filter is implemented in a typical sampled data system shown in Figure 18. 2. Note that there are several implicit requirements in the diagram. First, it is assumed that an ADC/DAC combination is available with sufficient sampling frequency, resolution, and dynamic range to accurately process the signal. Second, the DSP must be fast enough to complete all its calculations within the sampling interval, $1/f_s$. Third, analog filters are still required at the ADC imput and DAC output for anti-aliasing and anti-imaging, but the performance demands are not as great. Assuming these conditions have been met, the following offers a comparison between the digital and analog filters.

The required cutoff frequency of both filters is 1 kHz. The analog filter is realized as a 6-pole Chebyshev Type 1 filter (ripple in passband, no ripple in stopband), and the response is shown in Figure 18. 3. In practice, this filter would probably be realized using three 2-pole stages, each of which requires an op amp, and several resistors and capacitors. Modern filter design CAD packages make the 6-pole design relatively straightforward, but maintaining the 0. 5 dB ripple specification requires accurate component selection and matching.

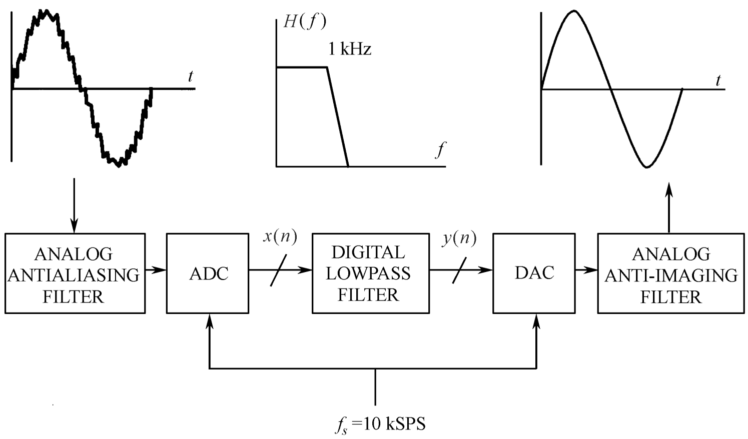


Figure 18. 2 A typical sampled data system

On the other hand, the 129-tap digital FIR filter shown has only 0. 002 dB passband ripple, linear phase, and a much sharper roll off. In fact, it could not be realized using analog techniques! Another obvious advantage is that the digital filter requires no component matching, and it is not sensitive to drift since the clock frequencies are crystal controlled. The 129-tap filter requires 129 MACs in order to compute an output sample. This processing must be completed within the sampling interval, $1/f_s$, in order to maintain real-time operation. In this example, the sampling frequency is 10 kSPS,

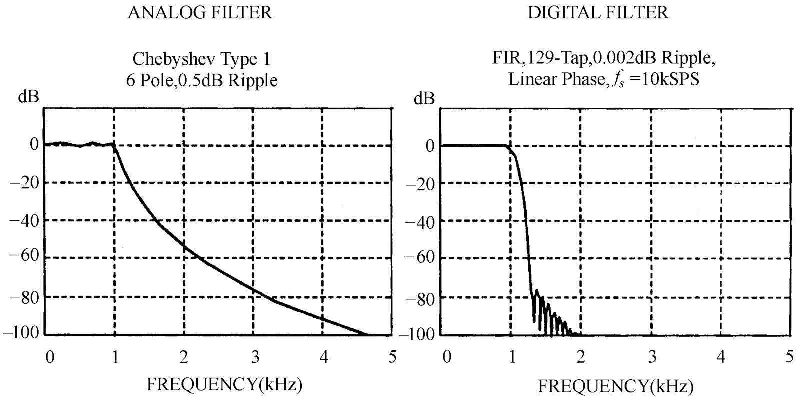


Figure 18.3 The magnitude response in dB

therefore 100 μs is available for processing, assuming no significant additional overhead requirement. Most DSPs can complete the entire multiply-accumulate process (and other functions necessary for the filter) in a single instruction cycle.

Therefore, a 129-tap filter requires that the instruction rate be greater than 129/100 μs=1.3 million instructions per second (MIPS). DSPs are available with instruction rates much greater than this, so the DSP certainly is not the limiting factor in this application.

In a practical application, there are certainly many other factors to consider when evaluating analog versus digital filters, or analog versus digital signal processing in general. Most modern signal processing systems use a combination of analog and digital techniques in order to accomplish the desired function and take advantage of the best of both the analog and the digital world.

New Words

clear-cut *adj*. 界限分明的
second-order *adj*. 二次的,二阶的
context [ˈkɔntekst] *n*. 上下文,环境,背景
desired [diˈzaiəd] *adj*. 期望的
tap [tæp] *n*. 抽头
Straightforward [ˌstreitˈfɔːwəd] *adj*. 简单易懂的
Overhead [ˈəuvəˈhed] *n*. 开销
Versus [ˈvəːsəs] *prəp*. 与……相比(可缩写为 vs. 或 v.)

Phrases & Expressions

with respect to 关于，至于

in the context of 在……情况下

at hand 在手边，在附近，即将到来

in reality 事实上，实际上

Technical Terms

filtering [ˈfiltəriŋ] *n*. 滤波

microphone [ˈmaikrəfəun] *n*. 麦克风

thermocouple [ˈθəːməuˌkʌpl] *n*. 热电偶

strain [strein] *n*. 张力

gage [geidʒ] *n*. 计量器

piezoelectric [paiˌiːzəuiˈlektrik] *adj*. 压电的

crystal [ˈkristl] *n*. 晶体

passband [ˈpɑːsbænd] *n*. 通带

stopband [ˈstɔpbænd] *n*. 阻带

package [ˈpækidʒ] *n*. 软件包，芯片封装

pole [pəul] *n*. 极点

ripple [ˈripl] *n*. 波纹

passive [ˈpæsiv] *adj*. 无源的

active [ˈæktiv] *adj*. 有源的

roll off *n*. 滚降

data compression 数据压缩

strain gage 张力计

piezoelectric crystal 压电晶体

disk drive head 磁盘驱动器磁头

signal conditioning circuit 信号调理电路

voltage reference 参考电压

cutoff frequency 截止频率

Chebyshev Type 1 filter 切比雪夫 1 型滤波器

FFT *abbr*. Fast Fourier transform 快速傅里叶变换

op amp *abbr*. operational amplifier 运算放大器

PCB *abbr*. Printed Circuit Board 印制电路板
CAD *abbr*. Computer Aided Design 计算机辅助设计
MAC *abbr*. Multiplication and Accumulation 乘法累加操作
SPS *abbr*. Sample Per Second 每秒样本数

Notes

1. 此句可译为：在一些情况下，选择（适当的）实现技术是很明确的；而在另外一些情况下，不存在非常明确的选择，可能需要考虑再三才能做出最终的决定。
2. 此句可译为：信号调理电路实际上就是模拟信号处理器。它们可以完成诸如放大（增益）、隔离（测量放大器和隔离放大器）、噪声环境下的信号检测（高共模抑制比的测量放大器）、动态范围压缩（对数放大器、对数数模转换器和增益可编程放大器）和滤波（有源滤波和无源滤波）之类的功能。
3. 此句可译为：模数转换器会以几种形式增加其复杂程度－ 提高采样率、扩大动态范围、提高精度、阻止输入噪声、对输入信号进行滤波、增益可编程放大器和片内参考电压等—— 所有这些都增加了模数转换器的功能从而简化了系统。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) What does "real-time" mean? In an ______(模拟系统), every task is performed in "real time" with ______(连续的)signals and processing. In a digital signal-processing (DSP) system, signals are represented with sets of samples, i. e. , values at ______(离散点)in time. Thus the time for processing a given number of samples in a DSP system can have an arbitrary interpretation in "real time", depending on the ______(采样率). In real-time applications, the sampling frequency must be at least ______ the frequency of the highest ______(频率分量)of interest in the analog signal. The time between samples ______(被称为)the sampling interval. To consider a system as operating in"real time", all processing of ______(一组给定的数据) (one or more samples, depending on the algorithm) must be completed ______(在新数据到来之前).

This definition of real time implies that, for a processor operating ______(以给定的时钟速率), the speed and quantity of the input data determines how much processing can be applied to the data without falling behind the data stream. The idea of having a limited amount of time with which to handle data may seem ______(奇怪的)to analog designers ______ this concept does not have a parallel in analog systems. In analog

systems, signals are processed continuously. The only penalty in a slow system is limited frequency response. By comparison, digital systems process parts of the signal, enough for very accurate approximations, but only within a limited block of time. Real-time DSP can be limited by ______(数据量)or type of processing that can be completed within the algorithm's time budget. For example, a given DSP processor handling data values sampled at, say, 48-kHz (audio signals), has less time to process those data values, including execution of all necessary tasks, than one sampling 8-kHz voice-band data.

(2) One can program a DSP to process data using one of several strategies for handling the"event," the arrival of data. A ______(状态位)or flag pin could be read periodically to determine whether new data is available. But"polling" wastes processor cycles. The data may arrive just after the last poll, but it can't make its presence known until the next poll. This ______(使实时系统开发变得困难).

The second strategy is for the data to interrupt the processor on arrival. Using interrupts to notify the processor is ______(有效率的), though not as easy to program; ______(时钟周期)can be wasted during the wait for an interrupt. Nevertheless, ______(事件驱动的中断编程)being well-suited to processing ______(现实世界信号)promptly, most DSPs are designed to deal efficiently with them. In fact, they are designed to ______(响应)very quickly to interrupts. In many DSP-based systems, the interrupt rates, based on the input data sampling rate, are often totally unrelated to the DSP's clock rate.

Because interrupt processing is such a vital element in DSP systems, processors typically have built-in hardware mechanisms to handle interrupts efficiently. Hard-wired mechanisms are more efficacious than software alone because a DSP's interrupt service routines (ISRs) may have to ______(满足以下所有要求):

- Fast context switching-switch from working on one task and its data (a context) to another context without the time loss and complication associated with writing programs to save register contents and chip status information.
- Nested-interrupt handling-handle ______(多个)interrupts of different priorities "simultaneously. "The DSP handles one interrupt at a time, but an interrupt of higher priority can take precedence over the handling of a lower-priority interrupt.
- Continue to accept data/record status—while the DSP services an interrupt, events keep on occurring in the real world and data keeps on arriving. To ______(跟上)the"real-world,"the DSP must record these events and accept the data—

then process them when it has finished servicing the interrupt.

In DSP systems, interrupts are typically generated by the arrival of data or the requirement to provide new output data. Interrupts may occur with each sample, or they may occur after ______(一帧数据)has been collected. The differences greatly influence how the DSP algorithm deals with data.

For algorithms that operate on a sample-by-sample basis, DSP software may be required to handle each incoming and outgoing data value. Each DSP ______(串行口) incorporates two data I/O registers, a receive register (Rx), and a transmit register (Tx). When a serial word is received, the port will typically generate a Receive interrupt. The processor ______(停下正在做的事情), begins executing code at the ______(中断向量) location, reads the incoming value from the Rx register into a processor data register, and either operates ______ that data value or returns to its background task.

______(为了发送数据), the serial port can generate a Transmit interrupt, indicating that new data can be written to the Tx register. The DSP can then begin code execution at the Tx interrupt vector and typically ______(转移一个数值)from a data register to the Tx register. If data input and output are controlled by the same sampling clock, only one interrupt is necessary. For example, if ______(一个程序段)is initiated by receive interrupt timing, new data would be read during the interrupt routine; then either the previously computed result, which is being held in a register, would be transmitted, ______ a new result would be computed and immediately transmitted—as the final step of the interrupt routine.

These mechanisms help a DSP to approach the ability ______ emulate what an analog system does naturally—continuously process data in real time—but with digital ______(精度和灵活性). In addition, in an efficiently programmed digital system, spare processor cycles left between processing data sets can be used to handle other tasks.

2. Translate the following passages into Chinese or English.

1) The primary reason for processing real-world signals is to extract information from them. This information normally exists in the form of signal amplitude (absolute or relative), frequency or spectral content, phase, or timing relationships with respect to other signals. Once the desired information is extracted from the signal, it may be used in a number of ways.

2) In some cases, it may be desirable to reformat the information contained in a signal. This would be the case in the transmission of a voice signal over a frequency

division multiple access (FDMA) telephone system. In this case, analog techniques are used to "stack" voice channels in the frequency spectrum for transmission via microwave relay, coaxial cable, or fiber. In the case of a digital transmission link, the analog voice information is first converted into digital using an ADC. The digital information representing the individual voice channels is multiplexed in time (time division multiple access, or TDMA) and transmitted over a serial digital transmission link (as in the T-Carrier system).

3) Another requirement for signal processing is to compress the frequency content of the signal (without losing significant information) then format and transmit the information at lower data rates, thereby achieving a reduction in required channel bandwidth. High speed modems and adaptive pulse code modulation systems (ADPCM) make extensive use of data reduction algorithms, as do digital mobile radio systems, MPEG recording and playback, and High Definition Television (HDTV).

4) The roots of DSP are in the 1960s and 1970s when digital computers first became available. Computers were expensive during this era, and DSP was limited to only a few critical applications. Pioneering efforts were made in four key areas: radar & sonar, where national security was at risk; oil exploration, where large amounts of money could be made; space exploration, where the data are irreplaceable; and medical imaging, where lives could be saved.

5) The first step in converting an analog signal into a digital signal is sampling. This step is accomplished using a sample and hold circuit, which grabs a sample from the signal and holds it steady until the next sampling time. Samples are normally collected at regular time intervals, called sampling periods. If the rate of the sampling is not adequate, distortion called aliasing can occur. The output from the sample and hold circuit is passed to an A/D converter, which chooses a quantization level that is closest to the actual amplitude of the signal. This is the second step in A/D conversion. The number of levels to choose from depends on the number of bits used by the converter: for N bits, 2^N levels are possible. Since analog amplitudes cannot in general be represented perfectly in a digital system, quantization errors occur.

6）数字信号处理(DSP)是无数家用和商用系统的关键部分,其应用领域与日俱增。因此,DSP 正在成为技术专家和工程师专业知识的重要组成部分之一。

7）信号(如声音、光和电压)是携带信息的变化。模拟信号是现实世界中的信号。模拟信号在每个时间点上都有定义,其幅度的取值可以是无限的。模拟信号不适合用计算机进行处理。经过取样和量化,模拟信号可以转换成为数字信号。数字信号仅仅在抽样点上有定义,其幅度仅能取有限的离散值。经过处理后,数字信号可以转换回模拟信号。

8）数字信号处理可分成“定点”和“浮点”两类。“定点”和“浮点”是指器件内数据的存储格式和运算格式。通常，定点数字信号处理器至少用16位来表示数据，而典型浮点数字信号处理器至少用32位来存储数值。

9）和其他科学和工程应用一样，数字信号处理器通常使用汇编语言或C语言编程。汇编语言程序执行起来快，C语言程序更易于开发和维护。在传统应用程序（如个人电脑上运行的程序）中，C语言几乎总是首选。如果真使用了汇编语言，那也只限于那些必须高速运行的、短小的子程序中。和传统软件执行的任务相比，DSP程序有两个重要的不同之处：(1)程序通常很短（比如，一百行而非一万行）；(2)执行速度总是DSP应用程序的关键。

Reading Materials

Passage 1 Typical DSP Applications

Telecommunications

Telecommunications is about transferring information from one location to another. This includes many forms of information: telephone conversations, television signals, computer files, and other types of data. To transfer the information, you need a channel between the two locations. This may be a wire pair, radio signal, optical fiber, etc. Telecommunications companies receive payment for transferring their customer's information, while they must pay to establish and maintain the channel. The financial bottom line is simple: the more information they can pass through a single channel, the more money they make. DSP has revolutionized the telecommunications industry in many areas: signalling tone generation and detection, frequency band shifting, filtering to remove power line hum, etc. Three specific examples from the telephone network will be discussed here: multiplexing, compression, and echo control.

Multiplexing

There are approximately one billion telephones in the world. At the press of a few buttons, switching networks allow any one of these to be connected to any other in only a few seconds. The immensity of this task is mind boggling! Until the 1960s, a connection between two telephones required passing the analog voice signals through mechanical switches and amplifiers. One connection required one pair of wires. In comparison, DSP converts audio signals into a stream of serial digital data. Since bits can be easily intertwined and later separated, many telephone conversations can be transmitted on a single channel. For example, a telephone standard known as the T-carrier system can simultaneously transmit 24 voice signals. Each voice signal is sampled 8000 times per second using an 8 bit companded (logarithmic compressed) analog-to-digital conversion. This results in each voice signal being represented as 64,000 bits/sec, and all 24 channels being contained in 1.544 megabits/sec. This signal can be transmitted about 6000 feet using ordinary telephone lines of 22 gauge copper wire, a typical interconnection distance. The financial advantage of digital transmission is enormous. Wire and analog switches are expensive; digital logic gates are cheap.

Compression

When a voice signal is digitized at 8000 samples/sec, most of the digital information

is redundant. That is, the information carried by any one sample is largely duplicated by the neighboring samples. Dozens of DSP algorithms have been developed to convert digitized voice signals into data streams that require fewer bits/sec. These are called data compression algorithms. Matching uncompression algorithms are used to restore the signal to its original form. These algorithms vary in the amount of compression achieved and the resulting sound quality. In general, reducing the data rate from 64 kilobits/sec to 32 kilobits/sec results in no loss of sound quality. When compressed to a data rate of 8 kilobits/sec, the sound is noticeably affected, but still usable for long distance telephone networks. The highest achievable compression is about 2 kilobits/sec, resulting in sound that is highly distorted, but usable for some applications such as military and undersea communications.

Echo control

Echoes are a serious problem in long distance telephone connections. When you speak into a telephone, a signal representing your voice travels to the connecting receiver, where a portion of it returns as an echo. If the connection is within a few hundred miles, the elapsed time for receiving the echo is only a few milliseconds. The human ear is accustomed to hearing echoes with these small time delays, and the connection sounds quite normal. As the distance becomes larger, the echo becomes increasingly noticeable and irritating. The delay can be several hundred milliseconds for intercontinental communications, and is particularly objectionable.

Digital Signal Processing attacks this type of problem by measuring the returned signal and generating an appropriate antisignal to cancel the offending echo. This same technique allows speakerphone users to hear and speak at the same time without fighting audio feedback (squealing). It can also be used to reduce environmental noise by cancelling it with digitally generated antinoise.

Audio Processing

The two principal human senses are vision and hearing. Correspondingly, much of DSP is related to image and audio processing. People listen to both music and speech. DSP has made revolutionary changes in both these areas.

Music

The path leading from the musician's microphone to the audiophile's speaker is remarkably long. Digital data representation is important to prevent the degradation commonly associated with analog storage and manipulation. This is very familiar to anyone who has compared the musical quality of cassette tapes with compact disks. In a

typical scenario, a musical piece is recorded in a sound studio on multiple channels or tracks. In some cases, this even involves recording individual instruments and singers separately. This is done to give the sound engineer greater flexibility in creating the final product. The complex process of combining the individual tracks into a final product is called mix down. DSP can provide several important functions during mix down, including: filtering, signal addition and subtraction, signal editing, etc. . One of the most interesting DSP applications in music preparation is artificial reverberation. If the individual channels are simply added together, the resulting piece sounds frail and diluted, much as if the musicians were playing outdoors. This is because listeners are greatly influenced by the echo or reverberation content of the music, which is usually minimized in the sound studio. DSP allows artificial echoes and reverberation to be added during mix down to simulate various ideal listening environments. Echoes with delays of a few hundred milliseconds give the impression of cathedral like locations. Adding echoes with delays of 10～20 milliseconds provide the perception of more modest size listening rooms.

Speech generation

Speech generation and recognition are used to communicate between humans and machines. Rather than using your hands and eyes, you use your mouth and ears. This is very convenient when your hands and eyes should be doing something else, such as: driving a car, performing surgery, or (unfortunately) firing your weapons at the enemy. Two approaches are used for computer generated speech: digital recording and vocal tract simulation. In digital recording, the voice of a human speaker is digitized and stored, usually in a compressed form. During playback, the stored data are uncompressed and converted back into an analog signal. An entire hour of recorded speech requires only about three megabytes of storage, well within the capabilities of even small computer systems. This is the most common method of digital speech generation used today. Vocal tract simulators are more complicated, trying to mimic the physical mechanisms by which humans create speech. The human vocal tract is an acoustic cavity with resonate frequencies determined by the size and shape of the chambers. Sound originates in the vocal tract in one of two basic ways, called voiced and fricative sounds. With voiced sounds, vocal cord vibration produces near periodic pulses of air into the vocal cavities. In comparison, fricative sounds originate from the noisy air turbulence at narrow constrictions, such as the teeth and lips. Vocal tract simulators operate by generating digital signals that resemble these two types of excitation. The characteristics of the resonate chamber are simulated by passing the excitation signal

through a digital filter with similar resonances. This approach was used in one of the very early DSP success stories, the Speak & Spell, a widely sold electronic learning aid for children.

Speech recognition

The automated recognition of human speech is immensely more difficult than speech generation. Speech recognition is a classic example of things that the human brain does well, but digital computers do poorly. Digital computers can store and recall vast amounts of data, perform mathematical calculations at blazing speeds, and do repetitive tasks without becoming bored or inefficient. Unfortunately, present day computers perform very poorly when faced with raw sensory data. Teaching a computer to send you a monthly electric bill is easy. Teaching the same computer to understand your voice is a major undertaking.

Digital Signal Processing generally approaches the problem of voice recognition in two steps: feature extraction followed by feature matching. Each word in the incoming audio signal is isolated and then analyzed to identify the type of excitation and resonate frequencies. These parameters are then compared with previous examples of spoken words to identify the closest match. Often, these systems are limited to only a few hundred words; can only accept speech with distinct pauses between words; and must be retrained for each individual speaker. While this is adequate for many commercial applications, these limitations are humbling when compared to the abilities of human hearing. There is a great deal of work to be done in this area, with tremendous financial rewards for those that produce successful commercial products.

Echo Location

A common method of obtaining information about a remote object is to bounce a wave off of it. For example, radar operates by transmitting pulses of radio waves, and examining the received signal for echoes from aircraft. In sonar, sound waves are transmitted through the water to detect submarines and other submerged objects. Geophysicists have long probed the earth by setting off explosions and listening for the echoes from deeply buried layers of rock. While these applications have a common thread, each has its own specific problems and needs. Digital Signal Processing has produced revolutionary changes in all three areas.

Radar

Radar is an acronym for RAdio Detection And Ranging. In the simplest radar system, a radio transmitter produces a pulse of radio frequency energy a few microseconds long. This

pulse is fed into a highly directional antenna, where the resulting radio wave propagates away at the speed of light. Aircraft in the path of this wave will reflect a small portion of the energy back toward a receiving antenna, situated near the transmission site. The distance to the object is calculated from the elapsed time between the transmitted pulse and the received echo. The direction to the object is found more simply; you know where you pointed the directional antenna when the echo was received.

The operating range of a radar system is determined by two parameters: how much energy is in the initial pulse, and the noise level of the radio receiver. Unfortunately, increasing the energy in the pulse usually requires making the pulse longer. In turn, the longer pulse reduces the accuracy and precision of the elapsed time measurement. This results in a conflict between two important parameters: the ability to detect objects at long range, and the ability to accurately determine an object's distance.

DSP has revolutionized radar in three areas, all of which relate to this basic problem. First, DSP can compress the pulse after it is received, providing better distance determination without reducing the operating range. Second, DSP can filter the received signal to decrease the noise. This increases the range, without degrading the distance determination. Third, DSP enables the rapid selection and generation of different pulse shapes and lengths. Among other things, this allows the pulse to be optimized for a particular detection problem. Now the impressive part: much of this is done at a sampling rate comparable to the radio frequency used, at high as several hundred megahertz! When it comes to radar, DSP is as much about high-speed hardware design as it is about algorithms.

Sonar

Sonar is an acronym for SOund NAvigation and Ranging. It is divided into two categories, active and passive. In active sonar, sound pulses between 2 kHz and 40 kHz are transmitted into the water, and the resulting echoes detected and analyzed. Uses of active sonar include: detection & localization of undersea bodies, navigation, communication, and mapping the sea floor. A maximum operating range of 10 to 100 kilometers is typical. In comparison, passive sonar simply listens to underwater sounds, which includes: natural turbulence, marine life, and mechanical sounds from submarines and surface vessels. Since passive sonar emits no energy, it is ideal for covert operations. You want to detect the other guy, without him detecting you. The most important application of passive sonar is in military surveillance systems that detect and track submarines. Passive sonar typically uses lower frequencies than active sonar because they propagate through the water with less absorption. Detection ranges can be

thousands of kilometers. DSP has revolutionized sonar in many of the same areas as radar: pulse generation, pulse compression, and filtering of detected signals. In one view, sonar is simpler than radar because of the lower frequencies involved. In another view, sonar is more difficult than radar because the environment is much less uniform and stable. Sonar systems usually employ extensive arrays of transmitting and receiving elements, rather than just a single channel. By properly controlling and mixing the signals in these many elements, the sonar system can steer the emitted pulse to the desired location and determine the direction that echoes are received from. To handle these multiple channels, sonar systems require the same massive DSP computing power as radar.

Reflection seismology

As early as the 1920s, geophysicists discovered that the structure of the earth's crust could be probed with sound. Prospectors could set off an explosion and record the echoes from boundary layers more than ten kilometers below the surface. These echo seismograms were interpreted by the raw eye to map the subsurface structure. The reflection seismic method rapidly became the primary method for locating petroleum and mineral deposits, and remains so today. In the ideal case, a sound pulse sent into the ground produces a single echo for each boundary layer the pulse passes through. Unfortunately, the situation is not usually this simple. Each echo returning to the surface must pass through all the other boundary layers above where it originated. This can result in the echo bouncing between layers, giving rise to echoes of echoes being detected at the surface. These secondary echoes can make the detected signal very complicated and difficult to interpret. Digital Signal Processing has been widely used since the 1960s to isolate the primary from the secondary echoes in reflection seismograms. How did the early geophysicists manage without DSP? The answer is simple: they looked in easy places, where multiple reflections were minimized. DSP allows oil to be found in difficult locations, such as under the ocean.

Image Processing

Images are signals with special characteristics. First, they are a measure of a parameter over space (distance), while most signals are a measure of a parameter over time. Second, they contain a great deal of information. For example, more than 10 megabytes can be required to store one second of television video. This is more than a thousand times greater than for a similar length voice signal. Third, the final judge of quality is often a subjective human evaluation, rather than an objective criteria. These

special characteristics have made image processing a distinct subgroup within DSP.

Medical

In 1895, Wilhelm Conrad Rötgen discovered that X-rays could pass through substantial amounts of matter. Medicine was revolutionized by the ability to look inside the living human body. Medical X-ray systems spread throughout the world in only a few years. In spite of its obvious success, medical X-ray imaging was limited by four problems until DSP and related techniques came along in the 1970s. First, overlapping structures in the body can hide behind each other. For example, portions of the heart might not be visible behind the ribs. Second, it is not always possible to distinguish between similar tissues. For example, it may be able to separate bone from soft tissue, but not distinguish a tumor from the liver. Third, X-ray images show anatomy, the body's structure, and not physiology, the body's operation. The X-ray image of a living person looks exactly like the X-ray image of a dead one! Fourth, X-ray exposure can cause cancer, requiring it to be used sparingly and only with proper justification.

The problem of overlapping structures was solved in 1971 with the introduction of the first computed tomography scanner (formerly called computed axial tomography, or CAT scanner). Computed tomography (CT) is a classic example of Digital Signal Processing. X-rays from many directions are passed through the section of the patient's body being examined. Instead of simply forming images with the detected X-rays, the signals are converted into digital data and stored in a computer. The information is then used to calculate images that appear to be slices through the body. These images show much greater detail than conventional techniques, allowing significantly better diagnosis and treatment. The impact of CT was nearly as large as the original introduction of X-ray imaging itself. Within only a few years, every major hospital in the world had access to a CT scanner. In 1979, two of CT's principle contributors, Godfrey N. Hounsfield and Allan M. Cormack, shared the Nobel Prize in Medicine. That's good DSP!

The last three X-ray problems have been solved by using penetrating energy other than X-rays, such as radio and sound waves. DSP plays a key role in all these techniques. For example, Magnetic Resonance Imaging (MRI) uses magnetic fields in conjunction with radio waves to probe the interior of the human body. Properly adjusting the strength and frequency of the fields cause the atomic nuclei in a localized region of the body to resonate between quantum energy states. This resonance results in the emission of a secondary radio wave, detected with an antenna placed near the body. The strength and other characteristics of this detected signal provide information about the localized region in resonance. Adjustment of the magnetic field allows the resonance region to be scanned throughout the body, mapping the

internal structure. This information is usually presented as images, just as in computed tomography. Besides providing excellent discrimination between different types of soft tissue, MRI can provide information about physiology, such as blood flow through arteries. MRI relies totally on Digital Signal Processing techniques, and could not be implemented without them.

Space

Sometimes, you just have to make the most out of a bad picture. This is frequently the case with images taken from unmanned satellites and space exploration vehicles. No one is going to send a repairman to Mars just to tweak the knobs on a camera! DSP can improve the quality of images taken under extremely unfavorable conditions in several ways: brightness and contrast adjustment, edge detection, noise reduction, focus adjustment, motion blur reduction, etc. Images that have spatial distortion, such as encountered when a flat image is taken of a spherical planet, can also be warped into a correct representation. Many individual images can also be combined into a single database, allowing the information to be displayed in unique ways. For example, a video sequence simulating an aerial flight over the surface of a distant planet.

Commercial Imaging Products

The large information content in images is a problem for systems sold in mass quantity to the general public. Commercial systems must be cheap, and this doesn't mesh well with large memories and high data transfer rates. One answer to this dilemma is image compression. Just as with voice signals, images contain a tremendous amount of redundant information, and can be run through algorithms that reduce the number of bits needed to represent them. Television and other moving pictures are especially suitable for compression, since most of the image remain the same from frame-to-frame. Commercial imaging products that take advantage of this technology include: video telephones, computer programs that display moving pictures, and digital television.

Questions:

1) Why the echo problems become serious as the distance of telephone connection becomes larger?

2) Are today's speech recognition systems satisfactory?

3) How DSP revolutionized radar system?

4) What problems used to limit medical X-ray imaging?

5) Can you tell some DSP applications in your daily life?

Passage 2　Software Radio

What Is a Software Radio?

The term software radio was coined by Joe Mitola in 1991 to refer to the class of reprogrammable or reconfigurable radios. In other words, the same piece of hardware can perform different functions at different times. The SDR Forum defines the ultimate software radio (USR) as a radio that accepts fully programmable traffic and control information and supports a broad range of frequencies, air-interfaces, and applications software. The user can switch from one air-interface format to another in milliseconds, use the GPS for location, store money using smartcard technology, or watch a local broadcast station or receive a satellite transmission.

The exact definition of a software radio is controversial, and no consensus exists about the level of reconfigurability needed to qualify a radio as a software radio. A radio that includes a microprocessor or DSP does not necessarily qualify as a software radio. However, a radio that defines in software its modulation, error correction, and encryption processes, exhibits some control over the RF hardware, and can be reprogrammed is clearly a software radio. A good working definition of a software radio is a radio that is substantially defined in software and whose physical layer behavior can be significantly altered through changes to its software. The degree of reconfigurability is largely determined by a complex interaction between a number of common issues in radio design, including systems engineering, antenna form factors, RF electronics, baseband processing, speed and reconfigurability of the hardware, and power supply management.

The term software radio generally refers to a radio that derives its flexibility through software while using a static hardware platform. On the other hand, a soft radio denotes a completely configurable radio that can be programmed in software to reconfigure the physical hardware. In other words, the same piece of hardware can be modified to perform different functions at different times, allowing the hardware to be specifically tailored to the application at hand. Nonetheless, the term software radio is sometimes used to encompass soft radios as well.

The functionality of conventional radio architectures is usually determined primarily by hardware with minimal configurability through software. The hardware consists of the amplifiers, filters, mixers (probably several stages), and oscillators. The software

is confined to controlling the interface with the network, stripping the headers and error correction codes from the data packets, and determining where the data packets need to be routed based on the header information. Because the hardware dominates the design, upgrading a conventional radio design essentially means completely abandoning the old design and starting over again. In upgrading a software radio design, the vast majority of the new content is software and the rest is improvements in hardware component design. In short, software radios represent a paradigm shift from fixed, hardware-intensive radios to multiband, multimode, software-intensive radios.

Characteristics and Benefits of a Software Radio

Implementation of the ideal software radio would require either the digitization at the antenna, allowing complete flexibility in the digital domain, or the design of a completely flexible RF front-end for handling a wide range of carrier frequencies and modulation formats. The ideal software radio, however, is not yet fully exploited in commercial systems due to technology limitations and cost considerations.

A model of a practical software radio is shown in Figure 1. The receiver begins with a smart antenna that provides a gain versus direction characteristic to minimize interference, multipath, and noise. The smart antenna provides similar benefits for the transmitter. Most practical software radios digitize the signal as early as possible in the receiver chain while keeping the signal in the digital domain and converting to the analog domain as late as possible for the transmitter using a DAC. Often the received signal is digitized in the intermediate frequency (IF) band. Conventional radio architectures employ a super heterodyne receiver, in which the RF signal is picked up by the antenna along with other spurious/unwanted signals, filtered, amplified with a low noise amplifier (LNA), and mixed with a local oscillator (LO) to an IF. Depending on the application, the number of stages of this operation may vary. Finally, the IF is then mixed exactly to baseband.

Digitizing the signal with an ADC in the IF range eliminates the last stage in the conventional model in which problems like carrier offset and imaging are encountered. When sampled, digital IF signals give spectral replicas that can be placed accurately near the baseband frequency, allowing frequency translation and digitization to be carried out simultaneously. Digital filtering (channelization) and sample rate conversion are often needed to interface the output of the ADC to the processing hardware to implement the receiver. Likewise, digital filtering and sample rate conversion are often necessary to interface the digital hardware that creates the modulated waveforms to the digital to

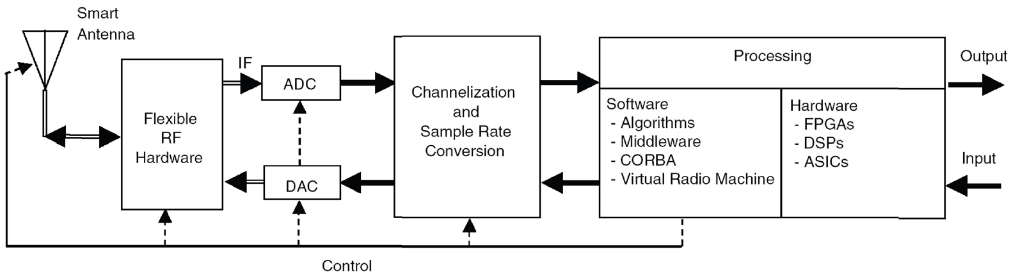


Figure 1 A model of a practical software radio

analog converter. Processing is performed in software using DSPs, FPGAs, or ASICs. The algorithm used to modulate and demodulate the signal may use a variety of software methodologies, such as middleware, e. g. , CORBA (common object request broker architecture), or virtual radio machines, which are similar in function to Java virtual machines. This forms a typical model of a software radio.

The software radio provides a flexible radio architecture that allows changing the radio personality, possibly in real-time, and in the process somewhat guarantees a desired QoS.

The flexibility in the architecture allows service providers to upgrade the infrastructure and market new services quickly. This flexibility in hardware architecture combined with flexibility in software architecture, through the implementation of techniques such as object-oriented programming and object brokers, provides software radio with the ability to seamlessly integrate itself into multiple networks with wildly different air and data interfaces. In addition, software radio architecture gives the system new capabilities that are easily implemented with software. For example, typical upgrades may include interference rejection techniques, encryption, voice recognition and compression, software-enabled power minimization and control, different addressing protocols, and advanced error recovery schemes.

Such capabilities are well-suited for 3G and 4G wireless requirements and advanced wireless networking approaches. In summary, five factors are expected to push wider acceptance of software radio.

1. **Multifunctionality**—With the development of short-range networks like Bluetooth and IEEE 802. 11, it is now possible to enhance the services of a radio by leveraging other devices that provide complementary services. For instance, a Bluetooth-enabled fax machine may be able to send a fax to a nearby laptop computer equipped with a software radio that supports the Bluetooth interface. Software radio's reconfiguration capability can support an almost infinite

variety of service capabilities in a system.

2. **Global mobility**—A number of communication standards exist today. In the 2G alone, there are IS-136, GSM, IS-95/CDMA1, and many other, less well known standards. The 3G technology tried to harmonize all the standards. However, there are many standards under the 3G umbrella. The need for transparency, i. e. , the ability of radios to operate with some, preferably all, of these standards in different geographical regions of the world has fostered the growth of the software radio concept. Military services also face a similar issue with incompatible radio standards existing between as well as within branches of the military.

3. **Compactness and power efficiency**—Multifunction, multimode radios designed using the "Velcro" approach of including separate silicon for each system can become bulky and inefficient as the number of systems increases. The software radio approach, however, results in a compact and, in some cases, a power-efficient design, especially as the number of systems increases, since the same piece of hardware is reused to implement multiple systems and interfaces.

4. **Ease of manufacture**—RF components are notoriously hard to standardize and may have varying performance characteristics. Optimization of the components in terms of performance may take a few years and thereby delay product introduction. In general, digitization of the signal early in the receiver chain can result in a design that incorporates significantly fewer parts, meaning a reduced inventory for the manufacturer.

5. **Ease of upgrades**—In the course of deployment, current services may need to be updated or new services may have to be introduced. Such enhancements have to be made without disrupting the operation of the current infrastructure. A flexible architecture allows for improvements and additional functionality without the expense of recalling all the units or replacing the user terminals. Vocoder technology, for example, is constantly improving to offer higher quality voice at lower bit rates. As new vocoders are developed, they can be quickly fielded in software radio systems. Furthermore, as new devices are integrated into existing infrastructures, software radio allows the new devices to interface seamlessly, from the air-interface all the way to the application, with the legacy network.

Users/Customers expect service regardless of the geographical areas in which they travel and the wireless technologies that are in use in different regions in the world, but carrying several devices that cover the broad range of technology alternatives is impractical. Users expect one device to utilize services in all regions, which is possible only by reconfiguring the receiver to the air-interface standards used in the respective regions. By dynamically

downloading the software to cover the needed air-interface standard, perhaps through transmission of the software configuration to the remote terminal, such over-the-air updates will allow for speedy implementation of software upgrades and new features.

Questions:

1) What does the concept of ***software radio*** refer to?
2) What does the term ***reconfigurability*** mean in this article?
3) What benefit do you think the most important that a software radio will bring?
4) Could you describe the functions of each block in the software radio model?
5) What factors will make software radio more widely accepted?

Passage 3　Digital Still Camera (DSC) System

Digital Photography Market Overview

Digital imaging, capturing, transferring, manipulation, and printing photos digitally came into its own in the early 90's thanks to the foundation laid a decade earlier by desktop publishing. Consumer-friendly printers, scanners, and digital cameras began to hit the market in 1993. Early versions of the Digital Still Camera (DSC) were expensive, poor in quality, and difficult to use. This scenario has changed in the past few years. Today, dozens of manufacturers produce a variety of compelling solutions. In 1997, the total number of digital cameras sold worldwide reached over 1 million units. This was due to several factors, such as the drop in PC prices, the explosion of the Internet and the ability to transfer and share pictures instantaneously, and the advances in VLSI technology.

Initially, digital photography was adopted in the commercial applications: badge printing, Web publishing, real estate, insurance companies, etc. Lately, however, this trend has changed dramatically and the casual home photographers clicking away at baby showers, birthday parties and wedding receptions are becoming the real mother lode. This is due to the drop in price of digital cameras, improved image quality, and a growing infrastructure. The DSC market has four major segments. Spatial resolution is the major classification for these segments. These segments are defined as follows:

Soft display (PC camera): These mobile cameras feature spatial resolution around 320×240, making them suitable primarily for soft display-based applications. They offer a limited set of camera features.

Basic point-and-shoot (low-end consumer): These units offer spatial resolution in the VGA range. They deliver a limited set of camera features, equivalent to an entry-level "point-and-shoot" film camera. In addition, various models offer one or more of the following premium features: zoom lens, autofocus, autoflash, removable storage, and color LCD screen. The resolution of these cameras ranges from 800 kpixels to 1.0 Mpixels.

Photo-quality point-and-shoot (high-end consumer): These cameras aim to provide the closest possible approximation to 35-mm film image quality at mass-market prices. The units in this category have resolution specifications ranging from 0.8 million to 3 million pixels.

Professional: These cameras are intended to replace 35-mm film cameras in professional news and documentation-gathering applications with high output quality requirements. They feature interchangeable lenses, standard high-end camera bodies and professional controls. Spatial resolution is in the range of 4～8 Mpixels.

DSC System

Digital still cameras require a significant amount of silicon contents, including the sensor (CCD or CMOS), the analogue components (ADC, NTSC encoder, …) and the engine (DSP), which is the brain of the camera and is responsible for performing all the computations needed to process and compress the image. Figure 1 shows the various functional blocks in a typical DSC system. Most DSCs use a CCD imager to sense the images. The driver electronics and the Timing Generator circuitry generate the necessary signal to clock the CCD. Correlated Double Sampling and Automatic Gain Control electronics are used to get a good-quality image signal from the CCD sensor. This CCD data is then digitised and fed into the DSC engine. All the image-processing and image-compression operations are performed in the DSC engine. On most DSCs, the user has the ability to view the image to be captured on the LCD display. The compressed images are stored in Flash memory for later use. Most DSC systems also provide an NTSC/PAL video signal to view the captured images (also the preview images) on a TV monitor. The current DSCs also provide ways to connect to the external PC or printer through an RS-232 or a USB port. Future DSC systems are expected to be even more versatile with the ability to annotate images with text/speech. Including a modem and TCP/IP interface provides the ability to connect directly to the Internet. Future DSCs will also run more complex multitasking operating systems to schedule the various real-time tasks.

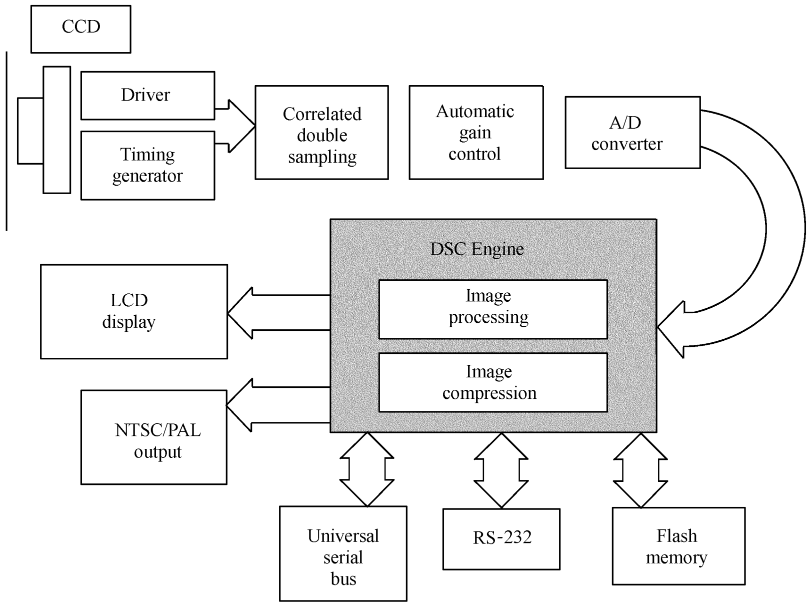


Figure 1　The DSC System

Image Acquisition

A typical DSC has to perform multiple processing steps before a high-quality image can be stored. The first step is image acquisition. The intensity distribution reflected from the scene is mapped by an optical system onto the imager. Nowadays, most cameras use CCDs although CMOS imagers are also used in some. The image captured by the CCD sensor has each pixel masked by a color filter to provide a color image. This raw CCD image is normally referred as a Color Filtered Array (CFA). The masking pattern of the CCD array, as well as the filter color primaries, vary among different manufacturers. In DSC applications, the CFA pattern that is most commonly used is an RGB Bayer pattern that consists of 2×2 cell elements that are tiled across the entire CCD-array. Figure 2 depicts a subset of this Bayer pattern in the matrix block following the CCD camera. The output signal of the CCD is digitised with a 10-bit or 12-bit A/D converter.

Image Pipeline

The CFA data needs to undergo significant amount of image processing before the image can be finally presented in a usable format for compression. All these processing stages are collectively called the"image pipeline". A typical image pipeline in a DSC is shown in Figure 2. As can be seen, a typical DSC has to perform multiple processing

steps before a high-quality image can be stored. Most of these tasks are multiply-accumulate (MAC) intensive operations. The TMS320C54xx DSP is well suited to perform these tasks efficiently and generate a high-quality image that is close to the image quality offered by traditional film from the raw CCD data. We outline the various image pipeline processing stages below.

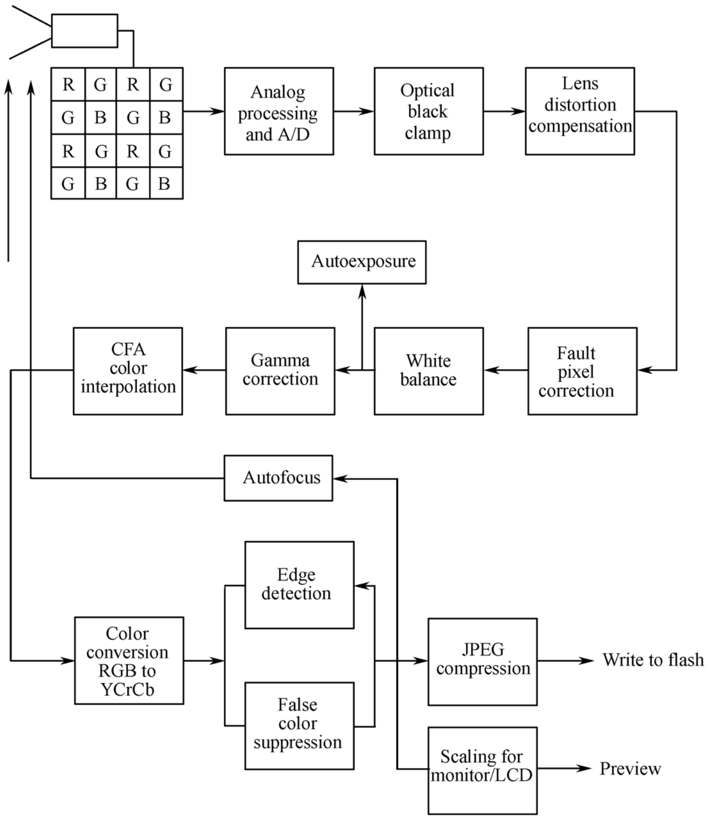


Figure 2 Image Pipeline

1. **Black Clamp** To optimise the dynamic range of the pixel values represented by the CCD imager, the pixels representing black need to be corrected since the CCD cell still records some non-zero current at these pixel locations. The black clamp function adjusts for this difference by subtracting an offset from each pixel value, but clamping/clipping to zero to avoid a negative result.

2. **Lens Distortion Compensation** Imperfections in the lens introduce nonlinearities in the brightness of the image. The image pipeline compensates for the lens by adjusting the brightness of each pixel, depending on its spatial location.

3. **Fault Pixel Interpolation** Large-pixel CCD arrays may have defective pixels. The manufacturer of the CCD sensor typically provides the locations of the missing pixels. The faulty pixel locations can also be computed by the DSC engine off-line by first capturing an image with the lens cap closed. The faulty pixels are imaged as "white spots" and the rest of the image is dark. The faulty pixel locations can then be identified with a simple threshold detector and stored in memory. During the normal operation of the DSC, the image values at the faulty pixel locations are computed by an interpolation technique.

4. **White Balance** The illumination during the recording of a scene is different from the illumination when viewing a picture. This results in a different color appearance that is typically seen as the bluish appearance of a face or the reddish appearance of the sky. Also, the sensitivity of each color channel varies such that grey or neutral colors are not represented correctly. To compensate for these unbalances in colors, the gain of the red, green, and blue channels is equalized. This is accomplished by computing the average brightness of each color component and by determining a scaling factor for each color component.

5. **CFA Interpolation** Due to the nature of a color filtered array, at any given pixel location, we only have one color pixel information (R, G or B in the case of a Bayer pattern). However, the image pipeline needs full color resolution (R, G and B) at each pixel in the image. Therefore, the two missing pixel colors are reconstructed by interpolating the neighbouring pixels.

6 . **Gamma Correction** Display devices used for image-viewing and printers used for image hardcopy have a nonlinear mapping between the image grey value and the actual displayed pixel intensities. The gamma correction stage compensates for the differences between the images generated by the CCD sensor and the image displayed on a monitor or printed into a page.

7. **Color Space Conversion** Typical image-compression algorithms such as JPEG operate on the YCbCr color space. Therefore, a color space conversion is performed to transform the image from an RGB color space to a YCbCr color space. This conversion is a linear transformation of each Y, Cb, and Cr value as a weighted sum of the R, G, B values at that pixel location.

8. **Edge Enhancement** The nature of CFA interpolation filters introduces a low-pass filter that smoothes the edges in the image. To sharpen the images, the image pipeline uses an edge detector to compute the edge magnitude in the Y channel at each pixel. The edge magnitude is then scaled and added to the original luminance (Y) image to enhance the sharpness of the image.

9. **False Color Suppression** Note that the edge enhancement is only performed in the Y channel of the image. This leads to misalignment in the color channels at the edges, resulting in rainbow-like artifacts. Suppressing the color components, Cb and Cr, at the edges reduces these artifacts.

10. **Autofocus** It is also possible to automatically adjust the lens focus in a DSC through image processing. These autofocus mechanisms operate in a feedback loop. They perform image processing to detect the quality of lens focus and move the lens motor iteratively until the image comes sharply into focus.

11. **Autoexposure** Due to varying scene brightness, to get a good overall image quality, it is necessary to control the exposure of the CCD. This is also accomplished in the DSC by sensing the average scene brightness and appropriately adjusting the CCD exposure time and/or gain. Similar to autofocus, the DSP performs this operation also in a closed-loop feedback fashion.

12. **Image Compression** Most DSCs are limited in the amount of memory available on the camera; hence, image compression is employed to reduce the memory requirements of captured images. Typically, compression ratios of about 10 : 1 to 15 : 1 are used. Most of the existing DSCs use JPEG compression. The DCT and Huffman encoding stages dominate the computations in JPEG. Future DSC will likely migrate to the JPEG2000 standard which employs a wavelet-coding scheme.

Questions:

1) Why digital photography was initially adopted in the commercial applications?
2) How is the DSC market divided according to this article?
3) What function does DSP perform in a DSC system?
4) How can a digital image file be obtained from the raw CCD data?
5) Why image compression is needed in DSCs?

Unit 7

Audio & Voice

Lesson 19　High Fidelity Audio

Audiophiles demand the utmost sound quality, and all other factors are treated as secondary. If you had to describe the mindset in one word, it would be: overkill [1]. Rather than just matching the abilities of the human ear, these systems are designed to exceed the limits of hearing [2]. It's the only way to be sure that the reproduced music is pristine [3]. Digital audio was brought to the world by the compact laser disc, or CD. This was a revolution in music; the sound quality of the CD system far exceeds older systems, such as records and tapes.

Figure 19.1 illustrates the surface of a compact laser disc, such as viewed through a high power microscope. The main surface is shiny (reflective of light), with the digital information stored as a series of dark pits burned on the surface with a laser. The information is arranged in a single track that spirals from the outside to the inside, the same as a phonograph record. The rotation of the CD is changed from about 210 to 480 rpm as the information is read from the outside to the inside of the spiral, making the scanning velocity a constant 1.2 meters per second. (In comparison, phonograph records spin at a fixed rate, such as 33, 45 or 78 rpm). During playback, an optical sensor detects if the surface is reflective or nonreflective, generating the corresponding binary information.

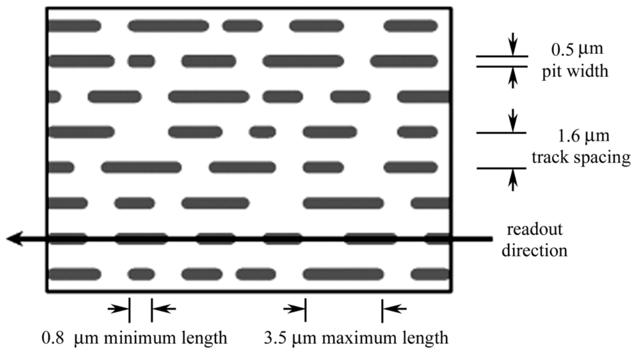


Figure 19.1　The surface of a compact disc

As shown by the geometry in Fig. 19.1, the CD stores about 1 bit per $(\mu m)^2$, corresponding to 1 million bits per $(mm)^2$, and 15 billion bits per disk. This is about

the same feature size used in integrated circuit manufacturing, and for a good reason. One of the properties of light is that it cannot be focused to smaller than about one-half wavelength, or 0.3 μm. Since both integrated circuits and laser disks are created by optical means, the fuzziness of light below 0.3 μm limits how small of features can be used.

Figure 19.2 shows a block diagram of a typical compact disc playback system. The raw data rate is 4.3 million bits per second, corresponding to 1 bit each 0.28 μm of track length. However, this is in conflict with the specified geometry of the CD; each pit must be no shorter than 0.8 μm, and no longer than 3.5 μm [4]. In other words, each binary one must be part of a group of 3 to 13 ones. This has the advantage of reducing the error rate due to the optical pickup, but how do you force the binary data to comply with this strange bunching?

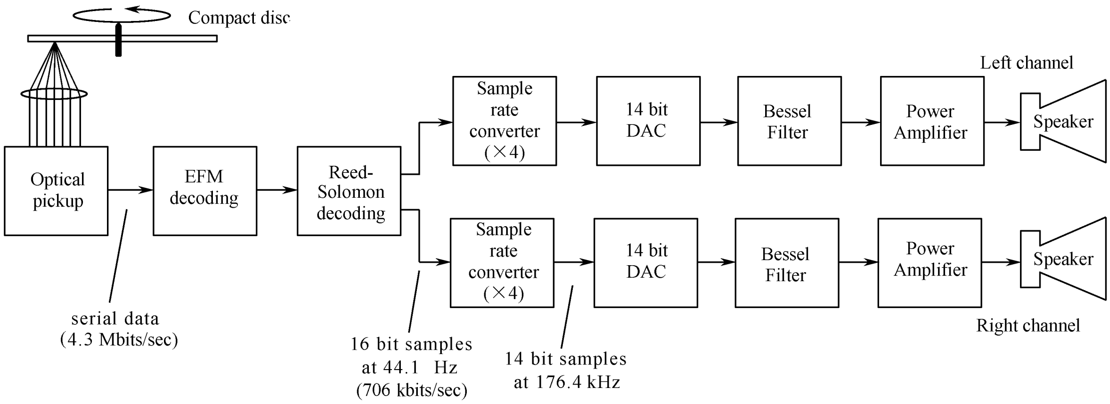


Figure 19.2 The block diagram of a typical compact disc playback system

The answer is an encoding scheme called eight-to-fourteen modulation (EFM). Instead of directly storing a byte of data on the disc, the 8 bits are passed through a look-up table that pops out 14-bits. These 14 bits have the desired bunching characteristics, and are stored on the laser disc. Upon playback, the binary values read from the disc are passed through the inverse of the EFM look-up table, resulting in each 14 bit group being turned back into the correct 8 bits.

In addition to EFM, the data are encoded in a format called two-level Reed-Solomon coding. This involves combining the left and right stereo channels along with data for error detection and correction. Digital errors detected during playback are either: corrected by using the redundant data in the encoding scheme, concealed by interpolating between adjacent samples, or muted by setting the sample value to

zero[5]. These encoding schemes result in the data rate being tripled, i. e., 1.4 Mbits/sec for the stereo audio signals versus 4.3 Mbits/sec stored on the disc.

After decoding and error correction, the audio signals are represented as 16 bit samples at a 44.1 kHz sampling rate. In the simplest system, these signals could be run through a 16 bit DAC, followed by a low-pass analog filter. However, this would require high performance analog electronics to pass frequencies below 20 kHz, while rejecting all frequencies above 22.05 kHz, 1/2 of the sampling rate[6]. A more common method is to use a multirate technique, that is, convert the digital data to a higher sampling rate before the DAC. A factor of four is commonly used, converting from 44.1 kHz to 176.4 kHz. This is called interpolation, and can be explained as a two step process (although it may not actually be carried out this way). First, three samples with a value of zero are placed between the original samples, producing the higher sampling rate. In the frequency domain, this has the effect of duplicating the 0 to 22.05 kHzspectrum three times, at 22.05 to 44.1 kHz, 44.1 to 66.15 kHz, and 66.15 to 88.2 kHz. In the second step, an efficient digital filter is used to remove the newly added frequencies.

The sample rate increase makes the sampling interval smaller, resulting in a smoother signal being generated by the DAC. The signal still contains frequencies between 20 Hz and 20 kHz; however, the Nyquist frequency has been increased by a factor of four. This means that the analog filter only needs to pass frequencies below 20 kHz, while blocking frequencies above 88. 2 kHz. This is usually done with a three pole Bessel filter. [7]

Since there are four times as many samples, the number of bits per sample can be reduced from 16 bits to 14 bits, without degrading the sound quality. The $\sin(x)/x$ correction[8] needed to compensate for the zero-order hold of the DAC can be part of either the analog or digital filter.

Audio systems with more than one channel are said to be in stereo (from the Greek word for solid, or three-dimensional). Multiple channels send sound to the listener from different directions, providing a more accurate reproduction of the original music. Music played through a monophonic (one channel) system often sounds artificial and bland. In comparison, a good stereo reproduction makes the listener feel as if the musicians are only a few feet away. Since the 1960s, high fidelity music has used two channels (left and right), while motion pictures have used four channels (left, right, center, and surround). In early stereo recordings, individual singers can often be heard in only one channel or the other. This rapidly progressed into a more sophisticated mix-down,

where the sound from many microphones in the recording studio is combined into the two channels. Mix-down is an art, aimed at providing the listener with the perception of being there.

The four channel sound used in motion pictures is called Dolby Stereo, with the home version called Dolby Surround Pro Logic. ("Dolby" and"Pro Logic"are trademarks of Dolby Laboratories Licensing Corp.)[9]. The four channels are encoded into the standard left and right channels, allowing regular two-channel stereo systems to reproduce the music. A Dolby decoder is used during playback to recreate the four channels of sound. The left and right channels, from speakers placed on each side of the movie or television screen, is similar to that of a regular two-channel stereo system (Figure 19.3). The speaker for the center channel is usually placed directly above or below the screen. Its purpose is to reproduce speech and other visually connected sounds, keeping them firmly centered on the screen, regardless of the seating position of the viewer/listener. The surround speakers are placed to the left and right of the listener, and may involve as many as twenty speakers in a large auditorium. The surround channel only contains midrange frequencies (say, 100 Hz to 7 kHz), and is delayed by 15 to 30 milliseconds. This delay makes the listener perceive that speech is coming from the screen, and not the sides. That is, the listener hears the speech coming from the front, followed by a delayed version of the speech coming from the sides. The listener's mind interprets the delayed signal as a reflection from the walls, and ignores it.

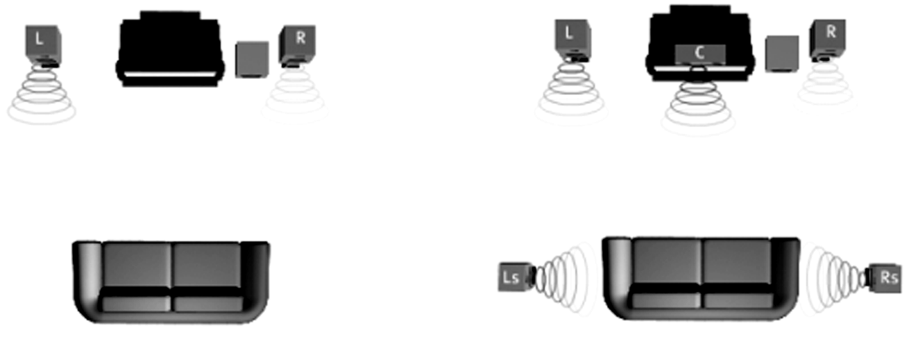


Figure 19.3 Two-channel Dolby Stereo (left) and 5.1-channel Dolby Pro Logic Ⅱ (right)

New Words

audiophile [ˈɔːdiəuˌfail] *n*. 高保真音响爱好者

utmost [ˈʌtməust] *n.* 极限 *adj.* 极度的
secondary [ˈsekəndəri] *adj.* 次要的
mindset [ˈmaindˌset] *n.* 观念,思维习惯
overkill [ˌəuvəˈkil] *n.* 超量杀伤力
hearing [ˈhiəriŋ] *n.* 听力,听觉
reproduce [ˌriːprəˈdjuːs] *v.* 繁殖,再生,重现
pristine [ˈpristin] *adj.* 质朴的,干净的,原始的
laser [ˈleizə] *n.* 激光
microscope [ˈmaikrəskəʊp] *n.* 显微镜
shiny [ˈʃaini] *adj.* 闪闪发光的
phonograph [ˈfəunəgrɑːf] *n.* 留声机,电唱机
reflective [riˈflektiv] *adj.* 反射的,反思的
pit [pit] *n.* 凹陷,坑
spiral [ˈspaiərəl] *n.* 螺旋线 *v.* 螺旋运动
focused [ˈfəukəst] *adj.* 聚焦的
fuzzy [ˈfʌzi] *adj.* 模糊的
data [ˈdeitə] *n.* 数据,资料
conflict [ˈkɔnflikt] *n.* 冲突,抵触
comply [kəmˈplai] *vi.* 遵守
inverse [inˈvəːs] *adj.* 相反的,逆向的
stereo [steriəu] *n.* 立体音响装置
mute [mjuːt] *v.* 使声音减弱或消失
triple [ˈtripl] *adj.* 三倍的
monophonic [ˌmɔnəˈfɔnik] *adj.* 单声部的,单声道的
artificial [ˌɑːtiˈfiʃl] *adj.* 不自然的
bland [blænd] *adj.* 平淡的
trademark [ˈtreidmɑːk] *n.* 商标
auditorium [ˌɔːdiˈtɔːriəm] *n.* 会堂,礼堂

Phrases & Expressions

in conflict with… 同……相冲突
comply with 同意,遵守

Technical Terms

playback [ˈpleibæk] *n.* 重放

encoding [inˈkəudiŋ] *n.* 编码

bunching [ˈbʌntʃiŋ] *n.* 聚束

scanning velocity 扫描速度

feature size 特征尺寸

block diagram 方框图

data rate 数据率

look-up table 查找表

Reed-Solomon coding 里德一索罗蒙编码

Bessel filter 贝塞耳滤波器

sin(x)/x correction sin(x)/x 校正

Dolby Stereo 杜比立体声

laser [ˈleizə] *abbr.* Light Amplification by Stimulated Emission of Radiation 激光

rpm *abbr.* revolutions per minute 转/分

Notes

1. 此句为虚拟语气,可译为:假如要用一个词来描述这种心理状态的话,那就是“过分”。
2. 此句可译为:这些系统不是设计成刚好满足人类听觉需求,而是超越了人类的听觉极限。
3. 此句可译为:这是确保再现音乐无任何失真的唯一方法。其中,“pristine”是指通过音响系统播放出来的音乐要和原来的现场音乐完全一样,没有任何失真和畸变。
4. 用作名词时,conflict 分别可以 with,of,between…and…连用,表示“和……的矛盾”、“……上的冲突”、“……和……之间的争执”,如 conflict of interest(利益冲突)。
5. 此句可译为:播放时检测到的错误,可以使用编码机制中的冗余数据进行纠正,也可以用邻近样本插值的方法去除,还可以用样本置零的方法使之“静音”。在本句中,使用了 either…,… or…结构连接了三个具有相同主语的从句。
6. 这两句隐含虚拟语气,因为实际中并没有采用所谓“最简系统”的方案。
7. 贝塞耳(Friedrich Wilhelm Bessel,1784—1846)是德国数学家、天文学家,他系统

总结了贝塞耳函数(该函数由 Daniel Bernoulli 发现)。贝塞耳滤波器是一类线性相位滤波器,其通带具备“恒定群延迟”特点,常用于高保真音响系统中。

8. 在数字信号处理和通信理论中,归一化 sinc 函数的定义为 sinc(x)=sinpx/px。在数学中,未归一化 sinc 函数的定义为 sinc(x)=sinx/x。
9. 杜比立体声(Dolby Stereo),又称杜比数码(Dolby Digital)。它是杜比(Dolby Laboratories, Inc.)公司开发的数字音频编解码系统,20 世纪 70 年代中期首先应用于电影院。杜比公司是用于消费类音视产品、娱乐媒体和专业音频的多声道音频技术的全球供应商。

Lesson 20 Audio Compression

CD-quality audio requires a transmission bandwidth of 1.411 Mbps. Clearly, substantial compression is needed to make transmission over the Internet practical. For this reason, various audio compression algorithms have been developed. Probably the most popular one is MPEG audio, which has three layers (variants), of which MP3 (MPEG audio layer 3) is the most powerful and best known. Large amounts of music in MP3 format are available on the Internet, not all of it legal, which has resulted in numerous lawsuits from the artists and copyright owners. MP3 belongs to the audio portion of the MPEG video compression standard.

Audio compression can be done in one of two ways. In waveform coding the signal is transformed mathematically by a Fourier transform into its frequency components. The amplitude of each component is then encoded in a minimal way. The goal is to reproduce the waveform accurately at the other end in as few bits as possible.

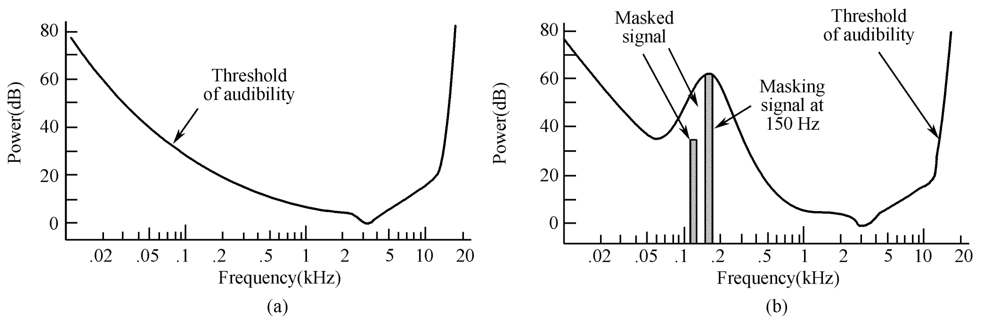
The other way, perceptual coding, exploits certain flaws in the human auditory system to encode a signal in such a way that it sounds the same to a human listener, even if it looks quite different on an oscilloscope. Perceptual coding is based on the science of psychoacoustics——how people perceive sound. MP3 is based on perceptual coding.

The key property of perceptual coding is that some sounds can mask other sounds. Imagine you are broadcasting a live flute concert on a warm summer day. Then all of a sudden, a crew of workmen nearby turn on their jackhammers and start tearing up the street. No one can hear the flute any more. Its sounds have been masked by the jackhammers. For transmission purposes, it is now sufficient to encode just the frequency band used by the jackhammers because the listeners cannot hear the flute anyway. This is called frequency

masking——the ability of a loud sound in one frequency band to hide a softer sound in another frequency band that would have been audible in the absence of the loud sound [1]. In fact, even after the jackhammers stop, the flute will be inaudible for a short period of time because the ear turns down its gain when they start and it takes a finite time to turn it up again. This effect is called temporal masking.

To make these effects more quantitative, imagine experiment 1. A person in a quiet room puts on headphones connected to a computer's sound card. The computer generates a pure sine wave at 100 Hz at low, but gradually increasing power. The person is instructed to strike a key when she hears the tone. The computer records the current power level and then repeats the experiment at 200 Hz, 300 Hz, and all the other frequencies up to the limit of human hearing. When averaged over many people, a log-log graph of how much power it takes for a tone to be audible looks like that of Figure 20.1(a). A direct consequence of this curve is that it is never necessary to encode any frequencies whose power falls below the threshold of audibility[2]. For example, if the power at 100 Hz were 20 dB in Figure 20.1(a), it could be omitted from the output with no perceptible loss of quality because 20 dB at 100 Hz falls below the level of audibility.

Now consider Experiment 2. The computer runs experiment 1 again, but this time with a constant-amplitude sine wave at, say, 150 Hz, superimposed on the test frequency [3]. What we discover is that the threshold of audibility for frequencies near 150 Hz is raised, as shown in Figure 20.1(b).



igure 20.1 (a) The threshold of audibility as a function of frequency. (b) The masking effect

The consequence of this new observation is that by keeping track of which signals are being masked by more powerful signals in nearby frequency bands, we can omit more and more frequencies in the encoded signal, saving bits [4]. In Figure20.1, the 125-Hz signal can be completely omitted from the output and no one will be able to hear

the difference. Even after a powerful signal stops in some frequency band, knowledge of its temporal masking properties allow us to continue to omit Figure 20. 1 (a). The threshold of audibility as a function of frequency. (b) The masking effect. the masked frequencies for some time interval as the ear recovers. The essence of MP3 is to Fourier-transform the sound to get the power at each frequency and then transmit only the unmasked frequencies, encoding these in as few bits as possible.

With this information as background, we can now see how the encoding is done. The audio compression is done by sampling the waveform at 32 kHz, 44. 1 kHz, or 48 kHz. Sampling can be done on one or two channels, in any of four configurations:

1. Monophonic (a single input stream).
2. Dual monophonic (e. g. , an English and a Japanese soundtrack).
3. Disjoint stereo (each channel compressed separately).
4. Joint stereo (interchannel redundancy fully exploited).

First, the output bit rate is chosen. MP3 can compress a stereo rock'n roll CD down to 96 kbps with little perceptible loss in quality, even for rock'n roll fans with no hearing loss. For a piano concert, at least 128 kbps are needed. These differ because the signal-to-noise ratio for rock's roll is much higher than for a piano concert (in an engineering sense, anyway). It is also possible to choose lower output rates and accept some loss in quality.

Then the samples are processed in groups of 1152 (about 26 msec worth). Each group is first passed through 32 digital filters to get 32 frequency bands. At the same time, the input is fed into a psychoacoustic model in order to determine the masked frequencies. Next, each of the 32 frequency bands is further transformed to provide a finer spectral resolution.

In the next phase the available bit budget is divided among the bands, with more bits allocated to the bands with the most unmasked spectral power, fewer bits allocated to unmasked bands with less spectral power, and no bits allocated to masked bands. Finally, the bits are encoded using Huffman encoding, which assigns short codes to numbers that appear frequently and long codes to those that occur infrequently.

There is actually more to the story. Various techniques are also used for noise reduction, antialiasing, and exploiting the interchannel redundancy, if possible, but these are beyond the scope of this book.

New Words

lawsuit [ˈlɔːsuːt] *n*. 诉讼
flaw [flɔː] *n*. 缺陷，瑕疵
psychoacoustics [ˌsaikəuəˈkuːstiks] *n*. 心理声学
live [laiv] *adj*. 实况转播的
flute [fluːt] *n*. 长笛
jackhammer [ˈdʒækˌhæmə] *n*. 手提钻
inaudible [inˈɔːdəbl] *adj*. 听不见的
quantitative [ˈkwɔntitətiv] *adj*. 数量的，定量的
superimpose [ˌsuːpərimˈpəuz] *v*. 叠加
disjoint [disˈdʒɔint] *adj*. 非结合的
perceptible [pəˈseptəbl] *adj*. 可察觉的，显而易见的
loss [lɔs] *n*. 损失
psychoacoustic [ˌsaikəuəˈkuːstik] *adj*. 心理声学的
budget [ˈbʌdʒit] *n*. 预算
allocate [ˈæləkeit] *vt*. 分配

Phrases & Expressions

all of a sudden 突然
in the absence of 缺乏……时，当……不在时
keep track of 记录

Technical Terms

oscilloscope [əˈsiləskəup] *n*. 示波器
mask [mɑːsk] *n*. 面具；掩模，掩码，掩蔽
threshold [ˈθreʃhəuld] *n*. 门限，阈值
monophonic [ˌmɔnəˈfɔnik] *adj*. 单声道的
redundancy [riˈdʌndənsi] *n*. 冗余
CD—quality audio CD 音质音频
transmission bandwidth 传输带宽
transmission over the Internet 互联网传输

waveform coding 波形编码
perceptual coding 感知编码
auditory system 听觉系统
frequency masking 频率掩蔽
temporal masking 暂时掩蔽
sound card 声卡
constant-amplitude 恒定幅度
spectral resolution 频谱分辨率
signal-to-noise ratio 信噪比
MPEG [ˈempeg] *abbr.* Moving Picture Experts Group 运动图像专家组

Notes

1. 此句可译为:这种现象称作"频率掩蔽"——某个频率上响度大的声音能够掩盖另一个频率上响度小的声音;如果没有响度大的声音,响度小的声音是可以听到的。在本句中,修饰 a softer sound in another frequency band 的定语从句 would have been audible in the absence of the loud sound 中含有虚拟语气。in the absence of the loud sound 是假设条件,而 would have been audible 是相应结果。
2. 此句可译为:根据这条曲线,可以直接得出如下结论——对功率在可听门限以下的频率成分进行编码是绝对没有必要的。
3. 此句可译为:计算机再次运行"实验 1",而这次使用固定幅度正弦波(比如 150Hz)叠加到测试频率上。with a constant-amplitude sine wave at, say, 150 Hz, superimposed on the test frequency 是由 with 引导的独立主格结构。
4. 此句可译为:通过这次观测,可以得到如下新成果——通过记录哪些信号会被邻近频带的强信号屏蔽,我们就可以把编码信号中更多的频率成分省略掉,从而节省了数据位数。

Lesson 21 Third-Generation Mobile Phones: Digital Voice and Data

What is the future of mobile telephony? Let us take a quick look. A number of factors are driving the industry. First, data traffic already exceeds voice traffic on the fixed network and is growing exponentially, whereas voice traffic is essentially flat. Many industry experts expect data traffic to dominate voice on mobile devices as well

soon [1]. Second, the telephone, entertainment, and computer industries have all gone digital and are rapidly converging. Many people are drooling over a lightweight, portable device that acts as a telephone, CD player, DVD player, e-mail terminal, Web interface, gaming machine, word processor, and more, all with worldwide wireless connectivity to the Internet at high bandwidth. This device and how to connect it is what third generation mobile telephony is all about.

Back in 1992, ITU tried to get a bit more specific about this dream and issued a blueprint for getting there called IMT-2000, where IMT stood for International Mobile Telecommunications. The number 2000 stood for three things: (1) the year it was supposed to go into service, (2) the frequency it was supposed to operate at (in MHz), and (3) the bandwidth the service should have (in kHz).

It did not make it on any of the three counts. Nothing was implemented by 2000. ITU recommended that all governments reserve spectrum at 2 GHz so devices could roam seamlessly from country to country[2]. Later, it was recognized that 2 Mbps is not currently feasible for users who are too mobile (due to the difficulty of performing handoffs quickly enough)[3]. More realistic is 2 Mbps for stationary indoor users (which will compete head-on with ADSL), 384 kbps for people walking, and 144 kbps for connections in cars.

The basic services that the IMT-2000 network is supposed to provide to its users are:

1. High-quality voice transmission.
2. Messaging (replacing e-mail, fax, SMS, chat, etc.).
3. Multimedia (playing music, viewing videos, films, television, etc.).
4. Internet access (Web surfing, including pages with audio and video).

Additional services might be video conferencing, telepresence[4], group game playing, and m-commerce. Furthermore, all these services are supposed to be available worldwide (with automatic connection via a satellite when no terrestrial network can be located), instantly (always on), and with quality-of-service guarantees.

ITU envisioned a single worldwide technology for IMT-2000, so that manufacturers could build a single device that could be sold and used anywhere in the world (like CD players and computers and unlike mobile phones and televisions). Having a single technology would also make life much simpler for network operators and would encourage more people to use the services. Format wars, such as the Betamax versus

VHS battle when videorecorders first came out, are not good for business.

Several proposals were made, and after some winnowing, it came down to two main ones. The first one, W-CDMA (Wideband CDMA), was proposed by Ericsson. This system uses direct sequence spread spectrum. It runs in a 5 MHz bandwidth and has been designed to interwork with GSM networks although it is not backward compatible with GSM. It does, however, have the property that a caller can leave a W-CDMA cell and enter a GSM cell without losing the call. This system was pushed hard by the European Union, which called it UMTS (Universal Mobile Telecommunications System).

The other contender was CDMA2000, proposed by Qualcomm. It, too, is a direct sequence spread spectrum design, basically an extension of IS-95 and backward compatible with it. It also uses a 5-MHz bandwidth, but it has not been designed to interwork with GSM and cannot hand off calls to a GSM cell . Other technical differences with W-CDMA include a different chip rate, different frame time, different spectrum used, and a different way to do time synchronization.

If the Ericsson and Qualcomm engineers were put in a room and told to come to a common design, they probably could. After all, the basic principle behind both systems is CDMA in a 5 MHz channel and nobody is willing to die for his preferred chip rate. The trouble is that the real problem is not engineering. Europe wanted a system that interworked with GSM; the U. S. wanted a system that was compatible with one already widely deployed in the U. S. (IS-95). Each side also supported its local company (Ericsson is based in Sweden; Qualcomm is in California). Finally, Ericsson and Qualcomm were involved in numerous lawsuits over their respective CDMA patents.

In March 1999, the two companies settled the lawsuits when Ericsson agreed to buy Qualcomm's infrastructure. They also agreed to a single 3G standard, but one with multiple incompatible options, which to a large extent just papers over the technical differences. These disputes notwithstanding[5], 3G devices and services are likely to start appearing in the coming years.

While waiting for the fighting over 3G to stop, some operators are gingerly taking a cautious small step in the direction of 3G by going to what is sometimes called 2.5G, although 2.1G might be more accurate. One such system is EDGE (Enhanced Data rates for GSM Evolution), which is just GSM with more bits per baud[6]. The trouble is, more bits per baud also means more errors per baud, so EDGE has nine different schemes for modulation and error correction, differing on how much of the bandwidth is devoted to fixing the errors introduced by the higher speed.

Another 2.5G scheme is GPRS (General Packet Radio Service), which is an overlay packet network on top of D-AMPS or GSM. It allows mobile stations to send and receive IP packets in a cell running a voice system. When GPRS is in operation, some time slots on some frequencies are reserved for packet traffic. The number and location of the time slots can be dynamically managed by the base station, depending on the ratio of voice to data traffic in the cell.

The available time slots are divided into several logical channels, used for different purposes. The base station determines which logical channels are mapped onto which time slots. One logical channel is for downloading packets from the base station to some mobile station, with each packet indicating who it is destined for[7]. To send an IP packet, a mobile station requests one or more time slots by sending a request to the base station. If the request arrives without damage, the base station announces the frequency and time slots allocated to the mobile for sending the packet. Once the packet has arrived at the base station, it is transferred to the Internet by a wired connection.

Nowadays, Some researchers are already working on 4G systems. Some of the proposed features of 4G systems include high bandwidth, ubiquity (connectivity everywhere), seamless integration with wired networks and especially IP, adaptive resource and spectrum management, software radios, and high quality of service for multimedia.

New Words

exponentially [ˌikspəuˈnenʃəli] *adv.* 以指数方式

dominate [ˈdɔmineit] *v.* 支配,占优势

drool [druːl] *vi.* 非常想要

reserve [riˈzəːv] *vt.* 预留 *n.* 储备

roam [rəum] *v.* 漫游,闲逛

seamlessly [ˈsiːmlisli] *adv.* 无缝地

head-on [ˈhedˌɔn] *adv.* 正面地

stationary [ˈsteiʃnri] *adj.* 固定的

fax [fæks] *n.* 传真

chat [tʃæt] *n.* 聊天

surfing [ˈsəːfiŋ] *n.* 冲浪

terrestrial [ti'restriəl] *adj*. 陆地的
envision [in'viʒən] *vt*. 想象,预想
winnow ['winəu] *vt*. 扬谷,精选,分出好坏
interwork [ˌintə'wəːk] *v*. 互相作用
push [puʃ] *vt*. 推动,推行
preferred [pri'fəːd] *adj*. 首选的
numerous ['njuːmərəs] *adj*. 许多的
patent ['peitnt] *n*. 专利
infrastructure [ˌinfrə'strʌktʃə] *n*. 基础设施
incompatible [ˌinkəm'pætəbl] *adj*. 不兼容的
notwithstanding [ˌnɔtwiθ'stændiŋ] *adv*. 虽然,尽管
gingerly ['dʒindʒəli] *adv*. 谨慎地,小心地
map [mæp] *v*. 映射
download ['daunˌləud] *v*. 下载
destined ['destind] *adj*. 去往……的,注定的
wired ['waiəd] *adj*. 有线的
deploy [di'plɔi] *v*. 部署,使用,配置
done [dʌn] *adj*. 结束的,已执行的,完成的
deal [diːl] *vi*. 处理,应付
ubiquity [juː'bikwəti] *n*. 到处存在,普遍存在
integration [ˌinti'ɡreiʃn] *n*. 综合,集成,积分

Phrases & Expressions

make it 达到预定目标,及时抵达
come down to 归结为,涉及
die for 非常需要
paper over 粉饰,掩盖
set up 设立,竖立

Technical Terms

handoff [ˈhændɔf] *n*. 越区切换
telepresence [ˈteliprezns] *n*. 遥现
m-commerce *n*. 移动商务
videorecorder [ˈvidiəu-riˈkɔːdə] *n*. 录像机
wideband [ˈwaidˈbænd] *adj*. 宽带的
synchronization [ˌsiŋkrənaiˈzeiʃn] *n*. 同步
baud [bɔːd] *n*. 波特
packet [ˈpækit] *n*. 信息包,分组
multimedia [ˈmʌltiˈmiːdjə] *n*. 多媒体
video conferencing 视频会议
network operator 网络运营商
backward compatible 向下兼容
chip rate 码片速率
IP packet IP 分组
time slot 时隙
logical channel 逻辑信道
base station 基站
software radio 软件无线电
ADSL *abbr*. Asymmetrical Digital Subscriber Loop 非对称数字用户线
SMS *abbr*. Short Message Service 短信业务
QoS *abbr*. Quality-of-Service 服务质量
VHS *abbr*. Video Home System 家用录像系统
DSSS *abbr*. Direct Sequence Spread Spectrum 直接序列扩频

Notes

1. 此句可译为:许多业界专家预期:在移动设备中,数据传输也将很快赶超语音业务。在本句中,dominate 是指数据传输业务将超过语音业务而在移动通信中占主导地位。
2. 此句可译为:国际电联建议各国政府预留 2G 的频谱,以便设备能够在国与国之间实现无缝漫游。
3. handoff (越区切换)是指当移动电话用户进入一个新的蜂窝区域时,信号传输将

从一个蜂窝切换到另一个蜂窝。这个切换过程在大约 0.25 秒内完成,这样通话者一般就不会注意到这种切换了。

4. Telepresence 是虚拟现实(Virtual Reality)技术的一种。该技术要创建一个多维信息构成的可操纵空间。虚拟现实技术最重要的目标就是实现真实体验和自然的人机交互,能够达到或者部分达到这样目标的系统就可称为虚拟现实系统。虚拟现实技术的其他形式还有"人工现实(Artificial Reality)"、"虚拟环境"(Virtual Environment)、"赛伯空间"(Cyberspace)等。
5. notwithstanding 可以用作介词、副词或连词。例如:The teams played on, notwithstanding the rain.(尽管下雨,各队仍然比赛。)又如:No matter how bad the weather is, the children will play foot-ball on the playground, notwithstanding.(不管天气怎么坏,孩子们还要在运动场踢足球。)
6. "波特"一词源自法国工程师 Jean Maurice Emile (1845—1903)。波特是一种数据传输速率单位,等于每秒内出现的离散状态或离散信号数。在二进制信号序列中,1 波特等于每秒 1 比特。
7. 此句可译为:一个逻辑信道用于从基站将分组下载到某个移动站,而每个分组都指明了自己要去谁那里。在本句中,destined 的意思是"去往……的"。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) The primary advantage of having two ears is the ability to ______(辨别声音的方向). Human listeners can detect the difference ______ two sound sources that are placed as little as three degrees apart, about the width of a person at 10 meters. This directional information is obtained ______ two separate ways. First, ______(约 1 kHz 以上的频率)are strongly shadowed by the head. In other words, the ear nearest the sound receives a stronger signal ______ the ear ______(在头的另一侧). The second clue to directionality is that the ear on the far side of the head hears the sound ______(比……稍微迟一些)the near ear, ______(由于它离声音源更远一些). ______(根据)a typical head size (about 22 cm) and the speed of sound (about 340 meters per second), an angular discrimination of three degrees requires a timing precision of about 30 microseconds. Since this timing requires the volley principle, this clue to directionality is predominately used for sounds less than about 1 kHz.

Both these sources of directional information are greatly aided by the ability to turn the head and observe the change in the signals. An interesting sensation occurs when a listener is presented ______ exactly the same sounds to both ears, such as listening to

monaural sound through headphones. The brain concludes that the sound is ______(来自听者头部的中央)!

______(尽管)human hearing can determine the direction a sound is ______, it does poorly in identifying the distance ______ the sound source. ______(这是因为)there are few clues available in a sound wave ______(可以提供这种信息). Human hearing weakly perceives that high frequency sounds are nearby, while low frequency sounds are distant. This is because sound waves dissipate their higher frequencies ______ they propagate long distances. Echo content is another weak clue to distance, providing a perception of the room size. For example, sounds in a large auditorium will contain echoes at about 100 millisecond intervals, while 10 milliseconds is typical for a small office. Some species have solved this ranging problem ______ using active sonar. For example, bats and dolphins produce clicks and squeaks that reflect ______ nearby objects. ______(通过测量发送信号和回波信号之间的时间间隔), these animals can locate objects ______ about 1 cm resolution. Experiments have shown that some humans, particularly the blind, can also use active ______(回声定位)to a small extent.

(2) In a CD (and any other digital recording technology), the goal is ______ create a recording with very high ______(保真度)(very high similarity ______ the original signal and the reproduced signal) and perfect reproduction (the recording sounds the same every single time you play it no matter how many times you play it).

______(为了实现这两个目标), digital recording converts the ______(模拟波形) into a stream of numbers and records the numbers instead of the wave. The conversion is done by ______(一种叫做模数转换器的器件). To play back the music, the stream of numbers is converted back to an analog wave ______ a digital-to-analog converter (DAC). The analog wave produced by the DAC ______ amplified and fed to the speakers to produce the sound.

The analog wave produced by the DAC will be the same every time, ______(只要) the numbers are not corrupted. The analog wave produced by the DAC will also ______(和……非常相似)the original analog wave if the analog-to-digital converter sampled ______ a high rate and produced accurate numbers.

You can understand why CDs have such high fidelity if you understand the analog-to-digital conversion process better. Let's say you have a sound wave, and you wish to sample it with an ADC. When you sample the wave with an analog-to-digital converter, you have control over two variables:

The sampling rate —Controls how many samples are taken per second.

The sampling precision—Controls how many different gradations (quantization

levels) are possible when taking the sample.

One thing about the CD's sampling rate and precision is ______ it produces a lot of data. On a CD, the digital numbers produced by the ADC ______(以字节形式存储), and it takes 2 bytes to represent 65,536 gradations. There are two sound streams being recorded (one for each of the speakers on a stereo system). ______(一张光盘可以存储多达74分钟的音乐), so the total amount of digital data that must be stored on a CD is: 44,100 samples/(channel × second) × 2 bytes/sample × 2 channels × 74 minutes × 60 seconds/minute = 783,216,000 bytes. That is a lot of bytes! To store that many bytes on a cheap piece of plastic that is tough enough to survive the abuse most people put a CD through is no small task, especially when you consider that the first CDs came out in 1980.

2. Translate the following passages into Chinese or English.

1) The range of human hearing is generally considered to be 20 Hz to 20 kHz, but it is far more sensitive to sounds between 1 kHz and 4 kHz. For example, listeners can detect sounds as low as 0 dB SPL at 3 kHz, but require 40 dB SPL at 100 hertz (an amplitude increase of 100). Listeners can tell that two tones are different if their frequencies differ by more than about 0.3% at 3 kHz. This increases to 3% at 100 hertz. For comparison, adjacent keys on a piano differ by about 6% in frequency.

2) The term octave means a factor of two in frequency. On the piano, one octave comprises eight white keys, accounting for the name (octo is Latin for eight). In other words, the piano's frequency doubles after every seven white keys, and the entire keyboard spans a little over seven octaves. The range of human hearing is generally quoted as 20 Hz to 20 kHz, corresponding to about 1/2 octave to the left, and two octaves to the right of the piano keyboard.

3) Companding can be carried out in three ways: (1) run the analog signal through a nonlinear circuit before reaching a linear 8 bit ADC, (2) use an 8 bit ADC that internally has unequally spaced steps, or (3) use a linear 12 bit ADC followed by a digital look-up table (12 bits in, 8 bits out). Each of these three options requires the same nonlinearity, just in a different place: an analog circuit, an ADC, or a digital circuit.

4) The perception of a continuous sound, such as a note from a musical instrument, is often divided into three parts: loudness, pitch, and timbre (pronounced "timber"). Loudness is a measure of sound wave intensity. Pitch is the frequency of the fundamental component in the sound. While there are subtle effects in both these perceptions, they are a straightforward

match with easily characterized physical quantities. Timbre is more complicated, being determined by the harmonic content of the signal.

5) Nearly all techniques for speech synthesis and recognition are based on the model of human speech production. Most human speech sounds can be classified as either voiced or fricative. Voiced sounds occur when air is forced from the lungs, through the vocal cords, and out of the mouth and/or nose. The vocal cords are two thin flaps of tissue stretched across the air flow, just behind the Adam's apple. In response to varying muscle tension, the vocal cords vibrate at frequencies between 50 and 1000 Hz, resulting in periodic puffs of air being injected into the throat. Vowels are an example of voiced sounds.

6）音频信号是一维声波信号。当声波进入耳朵，鼓膜振动，引起内耳小骨共振，将神经脉冲发送给大脑。人们把神经脉冲感知为声音。

7）音乐是一种非常重要的音频信号。语音是另一种重要的音频信号。人类语音通常在 600 Hz ～ 6000 Hz 范围内。语音由元音和辅音组成，两者具有不同的特性。

8）由于采用 16 比特样本和 44,100 样本/秒的采样率，一张音频 CD 所需的单声道带宽为 705.6 kbps，立体声带宽为 1.411 Mbps。这比视频所需带宽要小，但实时传输未经压缩的 CD 音质立体声信号仍要占用几乎整个 T1 信道。

9）电话系统中使用的“脉冲编码调制”采用了每秒 8000 个 8 比特样本。在北美和日本，7 个比特用于数据，1 个比特用于控制；在欧洲，8 个比特全部用于数据。“脉冲编码调制”系统的数据率为 56 000 bps 或 64 000 bps，采样率只有 8000 样本/秒，4 kHz 以上的频率就丢掉了。

10）计算机可以容易地利用软件对数字化声音进行处理。有几十种个人计算机程序允许用户对来自多种音源的声波进行录制、显示、编辑、混合和存储。目前，专业录音和编辑都采用数字技术。

Reading Materials

Passage 1　Sound Quality vs. Data Rate

When designing a digital audio system there are two questions that need to be asked: (1) how good does it need to sound? and (2) what data rate can be tolerated? The answer to these questions usually results in one of three categories. First, high fidelity music, where sound quality is of the greatest importance, and almost any data rate will be acceptable. Second, telephone communication, requiring natural sounding speech and a low data rate to reduce the system cost. Third, compressed speech, where reducing the data rate is very important and some unnaturalness in the sound quality can be tolerated. This includes military communication, cellular telephones, and digitally stored speech for voice mail and multimedia.

Table 1 shows the tradeoff between sound quality and data rate for these three categories. High fidelity music systems sample fast enough (44.1 kHz), and with enough precision (16 bits), that they can capture virtually all of the sounds that humans are capable of hearing. This magnificent sound quality comes at the price of a high data rate, 44.1 kHz×16 bits = 706 kbits/sec. This is pure brute force.

Table 1　The tradeoff between sould quality and data rate

Sound Quality Required	Bandwidth	Sampling rate	Number of bits	Datarate (Bits/sec)	Comments
High fidelity music (compact disc)	5 Hz to 20 kHz	44.1 kHz	16 bit	706k	Satisfies even the pickiest audiophile Better than human hearing
Telephone quality speech (with companding)	200 Hz to 3.2 kHz 200 Hz to 3.2 kHz	8 kHz 8kHz	12 bit 8 bit	96k 64k	Good speech quality; but very poor for music. Non-linear ADC reduces the data rate by 50%. A very common technique
Speech encoded by Linear Predictive Coding	200 Hz to 3.2 kHz	8 kHz	12 bit	4k	DSP speech compression technique. Very low data rates, poor voice quality

Whereas music requires a bandwidth of 20 kHz, natural sounding speech only

requires about 3.2 kHz. Even though the frequency range has been reduced to only 16% (3. 2 kHz out of 20 kHz), the signal still contains 80% of the original sound information (8 out of 10 octaves). Telecommunication systems typically operate with a sampling rate of about 8 kHz, allowing natural sounding speech, but greatly reduced music quality. You are probably already familiar with this difference in sound quality: FM radio stations broadcast with a bandwidth of almost 20 kHz, while AM radio stations are limited to about 3. 2 kHz. Voices sound normal on the AM stations, but the music is weak and unsatisfying.

Voice-only systems also reduce the precision from 16 bits to 12 bits per sample, with little noticeable change in the sound quality. This can be reduced to only 8 bits per sample if the quantization step size is made unequal. This is a widespread procedure called companding. An 8 kHz sampling rate, with an ADC precision of 8 bits per sample, results in a data rate of 64 kbits/sec. This is the brute force data rate for natural sounding speech. Notice that speech requires less than 10% of the data rate of high fidelity music.

The data rate of 64 kbits/sec represents the straightforward application of sampling and quantization theory to audio signals. Techniques for lowering the data rate further are based on compressing the data stream by removing the inherent redundancies in speech signals. One of the most efficient ways of compressing an audio signal is Linear Predictive Coding (LPC), of which there are several variations and subgroups. Depending on the speech quality required, LPC can reduce the data rate to as little as 2～6 kbits/sec.

Questions:

1) Does the sound quality of a digitised audio signal depend on its data rate?

2) What is data rate and how can it be figured out?

3) Which audio system has a letter sound quality, a CD player or a mobile phone?

4) Can you tell the reason why the quality of music is poor on AM stations while voices sound good?

5) Do you know any application of compressed speech?

Passage 2 Internet Radio

Once it became possible to stream audio over the Internet, commercial radio

stations got the idea of broadcasting their content over the Internet as well as over the air. Not so long after that, college radio stations started putting their signal out over the Internet. Then college students started their own radio stations. With current technology, virtually anyone can start a radio station. The whole area of Internet radio is very new and in a state of flux, but it is worth saying a little bit about.

There are two general approaches to Internet radio. In the first one, the programs are prerecorded and stored on disk. Listeners can connect to the radio station's archives and pull up any program and download it for listening. In fact, this is exactly the same as the streaming audio. It is also possible to store each program just after it is broadcast live, so the archive is only running, say, half an hour, or less behind the live feed. The advantages of this approach are that it is easy to do and listeners can pick and choose among all the programs in the archive.

The other approach is to broadcast live over the Internet. Some stations broadcast over the air and over the Internet simultaneously, but there are increasingly many radio stations that are Internet only. Some of the techniques that are applicable to streaming audio are also applicable to live Internet radio, but there are also some key differences.

One point that is the same is the need for buffering on the user side to smooth out jitter. By collecting 10 or 15 seconds worth of radio before starting the playback, the audio can be kept going smoothly even in the face of substantial jitter over the network. As long as all the packets arrive before they are needed, it does not matter when they arrived.

One key difference is that streaming audio can be pushed out at a rate greater than the playback rate since the receiver can stop it when the high-water mark is hit. Potentially, this gives it the time to retransmit lost packets, although this strategy is not commonly used. In contrast, live radio is always broadcast at exactly the rate it is generated and played back.

Another difference is that a live radio station usually has hundreds or thousands of simultaneous listeners whereas streaming audio is point to point. Under these circumstances, Internet radio should use multicasting with the RTP/RTSP protocols. This is clearly the most efficient way to operate. In current practice, Internet radio does not work like this. What actually happens is that the user establishes a TCP connection to the station and the feed is sent over the TCP connection. Of course, this creates various problems, such as the flow stopping when the window is full, lost packets timing out and being retransmitted, and so on.

The reason TCP unicasting is used instead of RTP multicasting is threefold. First,

few ISPs support multicasting, so that is not a practical option. Second, RTP is less well known than TCP and radio stations are often small and have little computer expertise, so it is just easier to use a protocol that is widely understood and supported by all software packages. Third, many people listen to Internet radio at work, which in practice, often means behind a firewall. Most system administrators configure their firewall to protect their LAN from unwelcome visitors. They usually allow TCP connections from remote port 25 (SMTP for email), UDP packets from remote port 53 (DNS), and TCP connections from remote port 80 (HTTP for the Web). Almost everything else may be blocked, including RTP. Thus, the only way to get the radio signal through the firewall is for the Web site to pretend it is an HTTP server, at least to the firewall, and use HTTP servers, which speak TCP. These severe measures, while providing only minimal security, often force multimedia applications into drastically less efficient modes of operation.

Since Internet radio is a new medium, format wars are in full bloom. RealAudio, Windows Media Audio, and MP3 are aggressively competing in this market to become the dominant format for Internet radio. A newcomer is Vorbis, which is technically similar to MP3 but open source and different enough that it does not use the patents MP3 is based on. A typical Internet radio station has a Web page listing its schedule, information about its DJs and announcers, and many ads. There are also one or more icons listing the audio formats it supports (or just LISTEN NOW if only one format is supported).

When a user clicks on one of the icons, the short metafile is sent over. The browser uses its MIME type or file extension to determine the appropriate helper (i. e. , media player) for the metafile. Then it writes the metafile to a scratch file on disk, starts the media player, and hands it the name of the scratch file. The media player reads the scratch file, sees the URL contained in it (usually with scheme http rather than rtsp to get around the firewall problem and because some popular multimedia applications work that way), contacts the server, and starts acting like a radio. As an aside, audio has only one stream, so http works, but for video, which has at least two streams, http fails and something like rtsp is really needed.

Another interesting development in the area of Internet radio is an arrangement in which anybody, even a student, can set up and operate a radio station. The main components are illustrated in Figure 1. The basis of the station is an ordinary PC with a sound card and a microphone. The software consists of a media player, such as Winamp or Freeamp, with a plug-in for audio capture and a codec for the selected output format,

for example, MP3 or Vorbis.

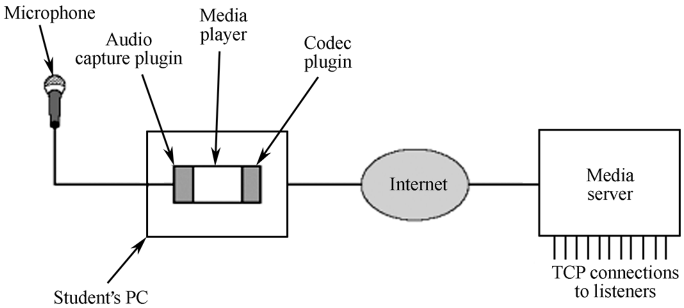


Figure 1　A student radio station

The audio stream generated by the station is then fed over the Internet to a large server, which handles distributing it to large numbers of TCP connections. The server typically supports many small stations. It also maintains a directory of what stations it has and what is currently on the air on each one. Potential listeners go to the server, select a station, and get a TCP feed. There are commercial software packages for managing all the pieces, as well as open source packages such as icecast. There are also servers that are willing to handle the distribution for a fee.

Questions:

1) Do you know something about streaming audio?
2) What differences exist between streaming audio and Internet radio?
3) Why is TCP unicasting currently used instead of RTP multicasting?
4) Can you describe how a typical Internet radio station works?
5) How does a student radio station work?

Passage 3　Voice-over IP (VoIP)

If you regularly make long-distance phone calls, chances are you've already used IP telephony without even knowing it. IP telephony, known in the industry as Voice-over IP (VoIP), is the transmission of telephone calls over a data network like one of the many networks that make up the Internet. While you probably have heard of VoIP, what you may not know is that many traditional telephone companies are already using it in the connections between their regional offices.

First, let's discuss the fundamental problem with existing telephone networks namely, their reliance on circuit switching. Circuit switching is a very basic concept that has been used by telephone networks for over 100 years. What happens is that when a call is made between two parties, the connection is maintained for the entire duration of the call. Because you are connecting two points in both directions, the connection is called a circuit. This is the foundation of the Public Switched Telephone Network (PSTN).

Here's how a typical telephone call works:

1) You pick up the receiver and listen for a dial tone. This lets you know that you have a connection to the local office of your telephone carrier.

2) You dial the number of the party you wish to talk to.

3) The call is routed through the switch at your local carrier to the party you are calling.

4) A connection is made between your telephone and the other party's line, opening the circuit.

5) You talk for a period of time and then hang up the receiver.

6) When you hang up, the circuit is closed, freeing your line.

Let's say that you talk for 10 minutes. During this time, the circuit is continuously open between the two phones. Telephone conversations over the traditional PSTN are transmitted at a fixed rate of about 64 kilobits per second (kbps) or 1 024 bits per second (bps), in each direction, for a total transmission rate of 128 kbps. Since there are 8 kilobits (kb) in a kilobyte (kB), this translates to a transmission of 16 kB each second the circuit is open, and 960 kB every minute it's open. So in a 10-minute conversation, the total transmission is 9600 kB, which is roughly equal to 9.4 megabytes (MB).

If you look at a typical phone conversation, much of this transmitted data is wasted. While you are talking, the other party is listening, which means that only half of the connection is in use at any given time. Based on that, we can surmise that we could cut the file in half, down to about 4.7 MB. Plus, a significant amount of the time in most conversations is dead air—for seconds at a time, neither party is talking. If we could remove these silent intervals, the file would be even smaller.

Data networks do not use circuit switching. Your Internet connection would be a lot slower if it maintained a constant connection to the Web page you were looking at. Instead of simply sending and retrieving data as you need it, the two computers involved in the connection would pass data back and forth the whole time, whether the data was

useful or not. That's no way to set up an efficient data network. Instead, data networks use a method called packet switching.

While circuit switching keeps the connection open and constant, packet switching opens the connection just long enough to send a small chunk of data, called a packet, from one system to another. What happens is this: The sending computer chops data into these small packets, with an address on each one telling the network where to send them. When the receiving computer gets the packets, it reassembles them into the original data.

Packet switching is very efficient. It minimizes the time that a connection is maintained between two systems, which reduces the load on the network. It also frees up the two computers communicating with each other so that they can accept information from other computers as well.

VoIP technology uses this packet-switching method to provide several advantages over circuit switching. For example, packet switching allows several telephone calls to occupy the amount of space occupied by only one in a circuit-switched network. Using PSTN, that 10-minute phone call consumed 10 full minutes of transmission time at a cost of 128 kbps. With VoIP, that same call may have occupied only 3.5 minutes of transmission time at a cost of 64 kbps, leaving another 64 kbps free for that 3.5 minutes, plus an additional 128 kbps for the remaining 6.5 minutes. Based on this simple estimate, another three or four calls could easily fit into the space used by a single call under the conventional system. And this example doesn't even factor in the use of data compression, which further reduces the size of each call.

Let's say that your company had equipment installed and a contract set up so that you can use VoIP. You have installed about a dozen telephones and a digital private branch exchange (PBX) in your office. A PBX is essentially a switch used to connect a number of phones (extensions) to each other and to one or more outside phone lines. In our example, the PBX is also a gateway.

Gateways are used to connect devices on two different types of networks so that they can communicate with each other. Our PBX is a gateway because it converts the standard circuit-switched signal from each phone into digital data that can be sent over a packet-switched, IP-based network. IP stands for Internet protocol, the language used by most data networks. Let's take another look at that typical telephone call, but this time using VoIP over a packet-switched network:

1) You pick up the receiver, which sends a signal to the PBX.

2) The PBX receives the signal and sends a dial tone. This lets you know that you

have a connection to the PBX.

3) You dial the number of the party you wish to talk to. This number is then temporarily stored by the PBX.

4) Once you have entered the number, the PBX checks it to ensure that it is in a valid format.

5) The PBX determines whom to map the number to. In mapping, the number is attached to the IP address of another device called the IP host. The IP host is typically another digital PBX that is connected directly to the phone system of the number you dialed. In some cases, particularly if the party you are calling is using a computer-based VoIP client, the IP host is the system you wish to connect with.

6) A session is established between your company's PBX and the other party's IP host. This means that each system knows to expect packets of data from the other system. Each system must use the same protocol to communicate. The systems will implement two channels, one for each direction, as part of the session.

7) You talk for a period of time. During the conversation, your company's PBX and the other party's IP host transmit packets back and forth when there is data to be sent. The PBX at your end keeps the circuit open between itself and your phone extension while it forwards packets to and from the IP host at the other end.

8) You finish talking and hang up the receiver.

9) When you hang up, the circuit is closed between your phone and the PBX, freeing your line.

10) The PBX sends a signal to the IP host of the party you called that it is terminating the session. The IP host terminates the session at its end, too.

11) Once the session is terminated, the PBX removes the number-to-IP-host mapping from memory.

Probably one of the most compelling advantages of packet switching is that data networks already understand the technology. By migrating to this technology, telephone networks immediately gain the ability to communicate the way computers do. Of course, having the ability to communicate and understanding the methods of communication are two very different things. For telephones to communicate with each other and with other devices, such as computers, over a data network, they need to speak a common language called a protocol.

There are two major protocols being used for VoIP. Both protocols define ways for devices to connect to each other using VoIP. Also, they include specifications for audio codecs. A codec, which stands for coder-decoder, converts an audio signal into a

compressed digital form for transmission and back into an uncompressed audio signal for replay.

The first protocol is H. 323, a standard created by the International Telecommunications Union (ITU). H. 323 is a comprehensive and very complex protocol (Table 1). It provides specifications for real-time, interactive videoconferencing, data sharing and audio applications such as IP telephony. Actually a suite of protocols, H. 323 incorporates many individual protocols that have been developed for specific applications.

Table 1 ITU-H. 323 Standard

H. 323 Protocol Suite			
Video	Audio	Data	Transport
H. 261 H. 263	G. 711 G. 722 G. 723. 1 G. 728 G. 729	T. 122 T. 124 T. 125 T. 126 T. 127	H. 225 H. 235 H. 245 H. 450. 1 H. 450. 2 H. 450. 3 RTP X. 224. 0

As you can see, full implementation of H. 323 requires a lot of overhead. Protocols. com: H. 323 provides detailed information about the entire H. 323 suite of protocols and how they relate to the OSI Reference Model.

An alternative to H. 323 emerged with the development of Session Initiation Protocol (SIP) under the auspices of the Internet Engineering Task Force (IETF). SIP is a much more streamlined protocol, developed specifically for IP telephony. Smaller and more efficient than H. 323, SIP takes advantage of existing protocols to handle certain parts of the process. For example, Media Gateway Control Protocol (MGCP) is used by SIP to establish a gateway connecting to the PSTN system. You can learn more about the architecture of SIP at Protocols. com: SIP.

There are four ways that you might talk to someone using VoIP. If you've got a computer or a telephone, you can use at least one of these methods without buying any new equipment:

Computer-to-computer—This is certainly the easiest way to use VoIP. You don't even have to pay for long-distance calls. There are several companies offering free or very low-cost software that you can use for this type of VoIP. All you need is the software, a microphone, speakers, a sound card and an Internet connection, preferably

a fast one like you would get through a cable or DSL modem. Except for your normal monthly ISP fee, there is usually no charge for computer-to-computer calls, no matter the distance.

Computer-to-telephone—This method allows you to call anyone (who has a phone) from your computer. Like computer-to-computer calling, it requires a software client. The software is typically free, but the calls may have a small per-minute charge.

Telephone-to-computer—A few companies are providing special numbers or calling cards that allow a standard telephone user to initiate a call to a computer user. The caveat is that the computer user must have the vendor's software installed and running on his or her computer. The good news is that the cost of the call is normally much cheaper than a traditional long-distance call.

Telephone-to-telephone—Through the use of gateways, you can connect directly with any other standard telephone in the world. To use the discounted services offered by several companies, you must call in to one of their gateways. Then, you enter the number you wish to call, and they connect you through their IP-based network. The downside is that you have to call a special number first. The upside is that the rates are typically much lower than standard long distance.

Although it will take some time to happen, you can be sure that, eventually, all of the circuit-switched networks will be replaced with packet-switching technology. IP telephony just makes sense, in terms of both economics and infrastructure requirements. More and more businesses are installing VoIP systems, and the technology will continue to grow in popularity as it makes its way into our homes.

Questions:

1) How does a typical telephone call work?
2) What're the differences between circuit switching and packet switching?
3) Do you know some existing VoIP standard?
4) What's your favorite way of using VoIP?
5) What's the future of VoIP?

Unit 8

Image & Video

Lesson 22 Digital Image Fundamentals

Digital Image Resolution

A digital image is made up of many rows and columns of pixels. For gray scale images, each pixel is assigned a number that represents the gray shade assigned to that pixel. The larger the number of pixels in an image, and the larger the number of available gray scale levels, the better the resolution of the image. The following image is an 8-bit images, with $2^8 = 256$ possible gray scale levels. The number of row and column are 808×562(Figure 22.1).

Figure 22.1 An 808×562×8 digital image

Histograms

The gray scales present in a digital image can be summarized by its histogram (see Figure 22.2). The histogram reports the number of pixels for each gray scale level present in the image, as a bar graph. When an image uses only a small portion of the available gray scale levels, histogram equalization can be used to spread out the usage of gray scale levels over the entire available range[1]. This procedure re-assigns gray scale levels so that image contrast is improved.

Addition and Subtraction of Images

Digital images can be added and subtracted pixel-by-pixel. Adding two images can

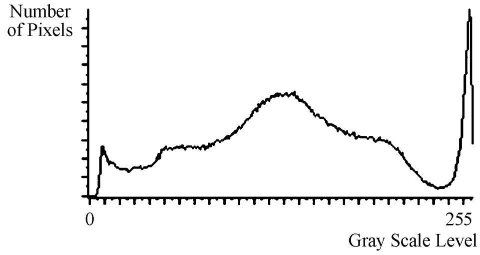


Figure 22.2　A histogram

combine two sets of objects into a single image. Moreover, several noisy images of the same scene can be averaged together to reduce the effect of noise. Image subtraction, on the other hand, can be used to remove an unwanted background from an image. Subtraction of two time-lapsed photographs will show where motion has occurred between the two.

When two images are added or subtracted, the resultant matrix will frequently contain illegal gray scale values. For example, when a pixel in one 8-bit image has the gray scale level 129 and the corresponding pixel in a second 8-bit image has the gray scale level 201, the sum pixel is 129+201=330. This is outside the legal range for an 8-bit image, which may only contain gray scale levels between 0 and 255. When the same two images are subtracted, the difference pixel is 129−201=−72, again a value outside the legal range. For these reasons, scaling follows most image arithmetic. Scaling to the range $[0,\ GSL_{max}]$ may be accomplished as follows:

$$\text{Scaled Output}[m,n]=\frac{\text{Output}[m,n]-\text{Min}(\text{Output})}{\text{Max}(\text{Output})-\text{Min}(\text{Output})}GSL_{max}$$

Warping and Morphing

Warping and morphing are digital image techniques that are finding application not only in entertainment but also in medical imaging. Warping stretches or re-shapes an object in an image, while morphing transforms one image into another. These transformations may be accomplished by marking control points, control lines, or triangles in a source image and choosing their new positions in a destination image. The transition between source and destination images is then accomplished by smoothly transforming not only the control element locations, but also their colors. The locations and colors of pixels not explicitly marked as control elements are determined by the locations and colors of the control elements that are nearest.

Image Filtering

Digital images can be filtered using two-dimensional convolution with a convolution kernel. When an $N \times N$ image is filtered by an $M \times M$ convolution kernel, $(M-1)/2$ rows and columns on each side of the image are lost to boundary effects. Low pass filters blur images, high pass filters emphasize sharp changes in gray scale level, and edge filters locate edges in an image[2].

Dilation and Erosion

Dilation adds a layer of pixels to all objects in an image. Erosion removes one layer of pixels from all objects. When dilation is followed by erosion, gaps in broken boundaries identified through edge detection can be filled in. Conversely, when erosion is followed by dilation, spots of noise in an image are removed.

Successfully detecting the edges in an image is the first step towards confident identification of object boundaries and then object recognition. From boundary information, shape characteristics like perimeter and area can be calculated, which can be used to classify an object.

Image Spectra

Two-dimensional FFTs are used to analyze the spectra of digital images. Just as in the one-dimensional case, a two-dimensional spectrum comprises a magnitude spectrum and a phase spectrum. The phase spectrum carries the best information about the locations of the objects in the image[3]. When all magnitudes are set to one, the phases alone still show a facsimile of the original image. When all phases are set to zero, the magnitudes alone show no trace of it.

Image spectra form the basis for both CT (computed tomography) and MRI (magnetic resonance imaging) scan displays. CT scans are X-rays taken in many directions in a single plane of an object [4]. MRI scans depend instead on the magnetic properties of an object placed in a varying magnetic field. Both types of scans permit non-invasive investigations of three-dimensional objects.

Image Compression

In part due to the Internet, digital images are transmitted from place to place more often than ever. To save time and bandwidth (space), both images and other files are often compressed before being transmitted. Lossless compression means that a file is

compacted without losing any information, so that the reconstructed file is identical to the original[5]. Lossy compression means that some information from the original file is irretrievably lost, but generally the reconstructed file is extremely close to the original. The compression ratio is the ratio of the original file size to the compressed file size.

One simple compression scheme is run-length encoding, which codes more than three repetitions of a number as three copies of the number followed by a count of the additional copies needed. Another compression scheme is Huffman encoding[6], which uses shorter codes to represent the most common signal elements, and longer codes to represent the least common signal elements.

JPEG, an extremely common image compression scheme, uses the discrete cosine transform (DCT) to concentrate most of the information about an 8×8 sub-block of an image into a few significant coefficients [7]. It then uses both run-length encoding and Huffman encoding to provide further compression.

New Words

shade [ʃeid] *n.* 色调明暗
arithmetic [əˈriθmətik] *n.* 算术
warping [ˈwɔːpiŋ] *n.* 扭曲,变形
morph [mɔf] *n.* 变体
dilation [daiˈleiʃn] *n.* 膨胀
erosion [iˈrəuʒən] *n.* 腐蚀
blur [bləː] *v.* 使……模糊
kernel [ˈkəːnl] *n.* 核心,内核
layer [ˈleiə] *n.* 层
identification [aiˌdentifiˈkeiʃn] *n.* 辨认,识别
perimeter [pəˈrimitə] *n.* 周长
facsimile [fækˈsiməli] *n.* 传真
lossless [lɔslis] *adj.* 无损的
lossy [ˈlɔsi] *adj.* 有损的

Phrases & Expressions

be made up of 由……组成
due to 由于

Technical Terms

pixel [ˈpiksəl] *n*. 像素

histogram [ˈhistəuɡræm] *n*. 直方图

gray scale image 灰度图像

gray scale level 灰度级

bar graph 条形图

histogram equalization 直方图均衡

image contrast 图像对比度

resultant matrix 结果矩阵

edge filter 边缘滤波器

edge detection 边缘检测

object recognition 目标识别

magnitude spectrum 幅度谱

phase spectrum 相位谱

Huffman encoding 哈夫曼编码

CT *abbr*. Computerized Tomography 计算机断层造影

MRI *abbr*. Magnetic Resonance Imaging 核磁共振成像

RLE *abbr*. Run－Length Encoding 行程编码

DCT *abbr*. Discrete Cosine Transform 离散余弦变换

Notes

1. 此句可译为：当一幅图像只用了可用灰度级的一小部分时，可以使用“直方图均衡”的方法将灰度级的使用扩展到整个可用范围。“直方图均衡”是一种提高图像对比度的方法。
2. 此句可译为：低通滤波器使图像变得模糊，高通滤波器突出了图像的灰度锐变，边缘滤波器对图像边缘进行定位。
3. 此句可译为：相位谱携带着图像中目标位置的信息。
4. 此句可译为：CT 就是用 X 射线从不同方向对目标的某个平面进行扫描。
5. 此句可译为：无损压缩是不损失任何信息地将文件进行压缩，重建文件和原始文件是完全相同的。
6. “哈夫曼编码”是一种用于无损数据压缩的熵编码算法，是 1952 年由哈夫曼(David A. Huffman)在其论文“A Method for the Construction of Minimum-

Redundancy Codes”中提出的。

7. 此句可译为：JPEG 是常用图像压缩方法之一，该方法使用“离散余弦变换”将图像中 8×8 小块的大部分信息集中到少数几个重要系数上。“离散余弦变换”是 Ahmed N.，T. Natarajan 和 K. R. Rao 于 1974 年提出的。1984 年，Chen W. H. 和 W. K. Pratt. 首先将其应用到图像压缩当中。

Lesson 23 Digital Camera

Just like a conventional camera, a digital camera has a series of lenses that focus light to create an image of a scene. But instead of focusing this light onto a piece of film, it focuses it onto a semiconductor device that records light electronically [1]. A microprocessor then breaks this electronic information down into digital data.

Image Sensors

The image sensor employed by most digital cameras is a charge coupled device (CCD). Some low-end cameras use complementary metal oxide semiconductor (CMOS) technology. While CMOS sensors will almost certainly improve and become more popular in the future, they probably won't replace CCD sensors in higher-end digital cameras.

CCDs use a special manufacturing process to create the ability to transport charge across the chip without distortion. This process leads to very high-quality sensors in terms of fidelity and light sensitivity [2]. CMOS chips, on the other hand, use completely standard manufacturing processes to create the chip - the same processes used to make most microprocessors. Because of the manufacturing differences, there are several noticeable differences between CCD and CMOS sensors.

1) CCD sensors create high-quality, low-noise images. CMOS sensors, traditionally, are more susceptible to noise.

2) CMOS sensors traditionally consume little power. CCDs, on the other hand, consume as much as 100 times more power than an equivalent CMOS sensor[3].

3) CMOS chips can be fabricated on just about any standard silicon production line, so they tend to be extremely inexpensive compared to CCD sensors. CCD sensors have been mass produced for a longer period of time, so they are more mature.

Resolution

Some typical resolutions that you find in digital cameras today include:

1) 640×480 pixels—This is the low end on most "real" cameras. This resolution is great if you plan to e-mail most of your pictures to friends or post them on a Web site. This is 307,000 total pixels.

2) 1216×912 pixels—If you are planning to print your images, this is a good resolution. This is a "megapixel" image size —1,109,000 total pixels.

3) 1600×1200 pixels—This is "high resolution." Images taken with this resolution can be printed in larger sizes, such as 8×10 inches, with good results. This is almost 2 million total pixels. You can find cameras today with up to 12 million pixels.

What picture resolution will give me the best quality prints on my inkjet printer? Kodak recommends the following as minimum resolutions for different print sizes (Table 23.1).

Table 23.1 Recommended minimum image resolutions for printing

Print Size	Megapixels	Image Resolution
Wallet	0.3	640×480 pixels
4×5 inches	0.4	768×512 pixels
5×7 inches	0.8	1152×768 pixels
8×10 inches	1.6	1536×1024 pixels

Capturing Color

In order to get a full color image, most sensors use filtering to look at the light in its three primary colors. Once all three colors have been recorded, they can be added together to create the full spectrum of colors.

There are several ways of recording the three colors in a digital camera. The highest quality cameras use three separate sensors, each with a different filter over it. Light is directed to the different sensors by placing a beam splitter in the camera. Think of the light entering the camera as water flowing through a pipe. Using a beam splitter would be like dividing an identical amount of water into three different pipes. Each sensor gets an identical look at the image; but because of the filters, each sensor only responds to one of the primary colors. The advantage of this method is that the camera records each of the three colors at each pixel location. Unfortunately, cameras that use this method tend to be bulky and expensive.

A second method is to rotate a series of red, blue and green filters in front of a single sensor. The sensor records three separate images in rapid succession. This method also provides information on all three colors at each pixel location; but since the three images aren't taken at precisely the same moment, both the camera and the target of the photo must remain stationary for all three readings. This isn't practical for candid photography or handheld cameras.

A more economical and practical way to record the three primary colors from a single image is to permanently place a filter over each individual photosite. By breaking up the sensor into a variety of red, blue and green pixels, it is possible to get enough information in the general vicinity of each sensor to make very accurate guesses about the true color at that location[4]. This process of looking at the other pixels in the neighborhood of a sensor and making an educated guess is called interpolation.

File formats

The two main file formats used by digital cameras are TIFF and JPEG. TIFF is an uncompressed format and JPEG is a compressed format. Most cameras use the JPEG file format for storing pictures, and they sometimes offer quality settings (such as medium or high). The following chart will give you an idea of the file sizes you might expect with different picture sizes(see Table 23.2).

Table 23.2 File formats and filesize

Image Size	TIFF(uncompressed)	JPEG(high quality)	JPEG(medium quality)
640×480	1.0MB	300kB	90kB
800×600	1.5MB	500kB	130kB
1024×768	2.5MB	800kB	200kB
1600×1200	6.0MB	1.7MB	420kB

Things you should keep in mind when purchasing a digital camera

Make sure the camera has the right resolution for your needs. If you are going to take snapshots and e-mail them to friends, then you don't need anything more than 640×480 pixel resolution. Buying the resolution that you need lets you save money (and hard disk space). On the other hand, if you want to print enlarged versions of your photos, you'll need a 2-megapixel or 3-megapixel camera.

Make sure the camera has enough memory. There is nothing more frustrating that "running out of film"when there is a great picture sitting in your viewfinder! The"film" for a digital camera is Flash memory, small hard disks, etc. Most cameras let you

download pictures from the camera so that you can take more, but if you go on a week-long vacation you will be away from your computer and won't be able to download. So make sure you pick up enough extra memory when you buy your camera so you won't run out when you need it. CompactFlash cards now come with up to 1 GB of space, so it's definitely possible to get all the memory you'll need for a long trip.

Make sure the lens will handle the pictures you plan to take. If you don't have the right lens, it can be hard to take the best pictures. For example, if very crisp detail is important in your pictures, you'll probably want a high optical zoom number. Be sure to try out the lens system on a camera before you purchase it. Digital cameras come with a huge variety of lenses, so be sure to shop around.

Do not confuse digital zoom with optical zoom. Many cameras advertise things like "100X zoom," but that is often misleading because only part of it is in the lens. The only part of a zoom lens that really matters is the "optical" part -- the part made out of glass lenses. This is the "zoom" that will increase the quality of the image. Any form of "digital zoom" is something you can do yourself outside of the camera. If you use your camera's software to crop out a small inner portion of a picture and blow it up, you are doing the same thing a digital zoom is doing. In most cases, the digital zoom simply makes the image fuzzy.

Do not confuse actual resolution with interpolated resolution. Many cameras advertise that they have, for example, 1000×600 pixel resolution and 1200×800 interpolated resolution. Like digital zoom, interpolated resolution is an illusion. You can do the same thing yourself with the camera's software, and all it really does is make the image larger and slightly fuzzy.

See how long the batteries will last. Many digital cameras eat batteries because they have to power an image sensor, an LCD panel and a microprocessor all at the same time, and sometimes there's a flash as well! See how long the batteries will really last in your camera. See if the camera will accept normal alkaline batteries in a pinch. If you plan on using your camera for long periods of time, think about purchasing an extra battery for it.

New Words

lens [lenz] *n.* 透镜,镜头
focus ['fəukəs] *n.* 焦点
susceptible [sə'septəbl] *adj.* 易受影响的

equivalent [iˈkwivələnt] *adj*. 相当的
filter [ˈfiltə] *n*. 滤光镜
bulky [ˈbʌlki] *adj*. 体积大的
succession [səkˈseʃn] *n*. 连续
photography [fəˈtɔgrəfi] *n*. 摄影
candid [ˈkændid] *adj*. 非排演的,偷拍的
hand-held *adj*. 手持的
vicinity [viˈsinəti] *n*. 邻近，附近
crisp [krisp] *adj*. 清晰的
viewfinder [ˈvjuːˌfaində] *n*. 取景器
crop [krɔp] *v*. 截切图像
zoom [zuːm] *n*. 缩放,变焦
panel [ˈpænl] *n*. 面板
interpolated [inˈtəːpəleitid] *adj*. 插值的
illusion [iˈluʒn] *n*. 错误观念;幻觉,错觉
alkaline [ˈælkəlain] *adj*. 碱性的

Phrases & Expressions

in succession 接连的
in a pinch 在紧急情况下
blow up 爆炸,放大

Technical Terms

sensor [ˈsensə] *n*. 传感器
distortion [disˈtɔːʃn] *n*. 畸变,失真
resolution [ˌrezəˈluːʃn] *n*. 分辨率
beam splitter 分光镜
CCD *abbr*. Charge Coupled Device 电荷耦合器件
TIFF *abbr*. Tagged Image File Format 标签图像文件格式
JPEG [ˈdʒeipeg] *abbr*. Joint Photographic Experts Group 联合图像专家组
LCD *abbr*. Liquid Crystal Display 液晶显示器

Notes

1. 此句可译为:数码相机不是将光线聚焦在胶片上,而是聚焦在半导体器件上,这种半导体器件采用电信号的形式来存储光信号。
2. 此句可译为:这种工艺使 CCD 传感器在"保真度"和"光敏性"方面具有非常好的质量。
3. 此句可译为:CCD 传感器的功耗比同等 CMOS 传感器多 100 倍。
4. 此句可译为:将传感器分成红、蓝、绿三组,就能在传感器的邻域获取足够信息,并对传感器所在位置的真实色彩做出比较准确的估计。

Lesson 24 Television Video Signals

Although over 50 years old, the standard television signal is still one of the most common way to transmit an image. Figure 24.1 shows how the television signal appears on an oscilloscope. This is called composite video, meaning that there are vertical and horizontal synchronization (sync) pulses mixed with the actual picture information[1]. These pulses are used in the television receiver to synchronize the vertical and horizontal deflection circuits to match the video being displayed. Each second of standard video contains 30 complete images, commonly called frames. A video engineer would say that each frame contains 525 lines, the television jargon for what programmers call rows. This number is a little deceptive because only 480 to 486 of these lines contain video information; the remaining 39 to 45 lines are reserved for sync pulses to keep the television's circuits synchronized with the video signal[2].

Standard television uses an interlaced format to reduce flicker in the displayed image. This means that all the odd lines of each frame are transmitted first, followed by the even lines. The group of odd lines is called the odd field, and the group of even lines is called the even field. Since each frame consists of two fields, the video signal transmits 60 fields per second. Each field starts with a complex series of vertical sync pulses lasting 1.3 milliseconds. This is followed by either the even or odd lines of video. Each line lasts for 63.5 microseconds, including a 10.2 microsecond horizontal sync pulse, separating one line from the next. Within each line, the analog voltage corresponds to the gray scale of the image, with brighter values being in the direction

away from the sync pulses. This place the sync pulses beyond the black range. In video jargon, the sync pulses are said to be blacker than black.

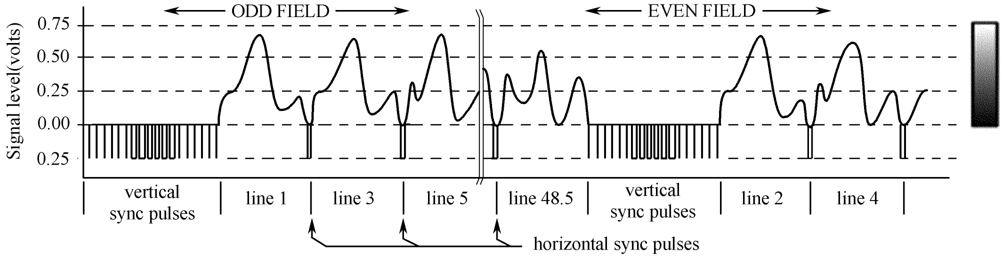


Figure 24.1 Composite video signal

The hardware used for analog-to-digital conversion of video signals is called a frame grabber. This is usually in the form of an electronics card that plugs into a computer, and connects to a camera through a coaxial cable. Upon command from software, the frame grabber waits for the beginning of the next frame, as indicated by the vertical sync pulses. During the following two fields, each line of video is sampled many times, typically 512, 640 or 720 samples per line, at 8 bits per sample. These samples are stored in memory as one row of the digital image.

This way of acquiring a digital image results in an important difference between the vertical and horizontal directions. Each row in the digital image corresponds to one line in the video signal, and therefore to one row of wells in the CCD. Unfortunately, the columns are not so straightforward. In the CCD, each row contains between about 400 and 800 wells (columns), depending on the particular device used. When a row of wells is read from the CCD, the resulting line of video is filtered into a smooth analog signal, such as in Figure 24.1. In other words, the video signal does not depend on how many columns are present in the CCD. The resolution in the horizontal direction is limited by how rapidly the analog signal is allowed to change. This is usually set at 3.2 MHz for color television, resulting in a rise time of about 100 nanoseconds, i.e., about 1/500th of the 53.2 microsecond video line.

When the video signal is digitized in the frame grabber, it is converted back into columns. However, these columns in the digitized image have no relation to the columns in the CCD. The number of columns in the digital image depends solely on how many times the frame grabber[3] samples each line of video. For example, a CCD might have 800 wells per row, while the digitised image might only have 512 pixels (i.e., columns) per row.

The number of columns in the digitized image is also important for another reason. The standard television image has an aspect ratio of 4 to 3, i. e. , it is slightly wider than it is high. Motion pictures have the wider aspect ratio of 25 to 9. CCDs used for scientific applications often have an aspect ratio of 1 to 1, i. e. , a perfect square. In any event, the aspect ratio of a CCD is fixed by the placement of the electrodes, and cannot be altered. However, the aspect ratio of the digitized image depends on the number of samples per line. This becomes a problem when the image is displayed, either on a video monitor or in a hardcopy[4]. If the aspect ratio isn't properly reproduced, the image looks squashed horizontally or vertically.

The 525 line video signal described here is called NTSC (National Television Systems Committee)[5], a standard defined way back in 1954. This is the system used in the United States and Japan. In Europe there are two similar standards called PAL (Phase Alternation by Line)[6] and SECAM (Sequential Chrominance And Memory)[7]. The basic concepts are the same, just the numbers are different. Both PAL and SECAM operate with 25 interlaced frames per second, with 625 lines per frame. Just as with NTSC, some of these lines occur during the vertical sync, resulting in about 576 lines that carry picture information. Other more subtle differences relate to how color and sound are added to the signal.

The most straightforward way of transmitting color television would be to have three separate analog signals, one for each of the three colors the human eye can detect: red, green and blue[8]. Unfortunately, the historical development of television did not allow such a simple scheme. The color television signal was developed to allow existing black and white television sets to remain in use without modification. This was done by retaining the same signal for brightness information, but adding a separate signal for color information. In video jargon, the brightness is called the luminance signal, while the color is the chrominance signal. The chrominance signal is contained on a 3. 58 MHz carrier wave added to the black and white video signal. Sound is added in this same way, on a 4. 5 MHz carrier wave. The television receiver separates these three signals, processes them individually, and recombines them in the final display.

New Words

deflection [di'flekʃn] *n.* 偏转

frame [freim] *n.* 帧

jargon ['dʒɑːgən] *n.* 行话

deceptive [di'septiv] *adj*. 欺骗性的

interlace [ˌintə'leis] *vt*. 交织，交错

flicker ['flikə] *n*. 闪烁，颤动

squash [skwɔʃ] *v*. 挤进，挤压

Phrases & Expressions

in any event 无论如何

Technical Terms

luminance ['luːminəns] *n*. 亮度

chrominance ['krəuminəns] *n*. 色度

carrier wave 载波

odd field 奇数场

even field 偶数场

frame grabber 帧采集器

NTSC *abbr*. National Television Systems Committee 国家电视系统委员会

PAL [pæl] *abbr*. Phase Alternation by Line 逐行倒相

SECAM ['siːkæm] *abbr*. SEquential Couleur Avec Memoire 顺序与存储彩色电视系统

Notes

1. 此句可译为：该信号被称为“复合视频信号”，意思是该信号在图像信号中加入了垂直和水平同步脉冲。
2. 此句可译为：这个数字有点名不副实，因为只有480～486线包含视频信息，而剩下的29～45线用作视频信号同步。
3. “帧采集器”(frame grabber)是计算机系统中用于模拟视频信号数字化的设备，其功能和显卡正好相反。
4. 此句可译为：在视频监视器上显示数字图像时或进行数字图像硬拷贝时，这就会成为问题。
5. NTSC制式(又称N制)是1952年12月由美国国家电视系统委员会(NTSC)制定的彩色电视广播标准。
6. PAL制式(帕尔制)是1962年由前联邦德国在NTSC技术上研制出来的改进方

案。

7. SECAM 制式(塞康制)是 1966 年由法国研制成功的。

8. 此句可译为:传送彩色电视信号最直接的方法可能是用三路独立模拟信号,分别代表人眼可以检测到的红、绿、蓝三种颜色之一。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) ______(为了理解视频), it is best to start with simple, old-fashioned black-and-white television. To represent ______(二维图像)in front of it as a one-dimensional voltage as ______(时间的函数), the camera scans an ______(电子束)rapidly across the image and slowly down it, recording ______(光的强度)as it goes. At the end of the scan, ______(称作一帧), the beam retraces. This intensity as a function of time is broadcast, and receivers repeat the scanning process to ______(重建图像).

The exact scanning parameters ______ from country to country. The system used in North and South America and Japan has 525 ______(扫描线), a horizontal-to-vertical aspect ratio of 4∶3, and 30 frames/sec. The European system has 625 scan lines, the same aspect ratio of 4∶3, and 25 frames/sec. ______(在这两个系统中), the top few and bottom few lines are not displayed (to approximate a rectangular image on the original round CRTs). Only 483 of the 525 NTSC scan lines (and 576 of the 625 PAL/SECAM scan lines) are displayed. The beam ______(被关闭)during the vertical retrace, so many stations (especially in Europe) use this time to broadcast TeleText (text pages containing news, weather, sports, stock prices, etc.).

While 25 frames/sec is enough ______ capture smooth motion, at that frame rate many people, especially older ones, will perceive the image to flicker (because the old image ______ faded off the retina before the new one appears). ______ increase the frame rate, which would require using more scarce bandwidth, a different approach is taken. ______ the scan lines being displayed ______(依次), first all the odd scan lines are displayed, then the even ones are displayed. Each of these half frames is called a field. Experiments have shown ______ although people notice flicker at 25 frames/sec, they do not notice it at 50 fields/sec. This technique is called interlacing. Noninterlaced television or video is called progressive. Note that movies run ______ 24 fps, but each frame is fully visible for 1/24 sec.

(2) Color video uses the same scanning pattern ______ monochrome (black and white), except ______ instead of displaying the image with one moving beam, it uses

three beams moving ______ unison. One beam is used ______ each of ______（加性三基色）: red, green, and blue (RGB). This technique works because any color can be constructed from ______（红、绿、蓝的线性叠加）with the appropriate intensities. However, for transmission on a single channel, the three color signals ______（必须合成为单一复合信号）.

When color television was invented, ______（各种用来显示色彩的方法）were technically possible, and different countries ______ made different choices, leading to systems that are still ______（不兼容的）.(Note that these choices have nothing to do with VHS versus Betamax versus P2000, which are recording methods.) In all countries, a political requirement was that programs transmitted in color had to be receivable on ______（存在的）black-and-white television sets. Consequently, the simplest scheme, just encoding the RGB signals separately, was not acceptable. ______（RGB 也不是效率最高的方案）.

2. Translate the following passages into Chinese or English.

1) The three-dimensional world is imaged by the lens of the human eye onto the retina, which is populated with photoreceptor cells that respond to light having wavelengths in the range of about 400 nm to 700 nm. In an imaging system, we build a camera having a lens and a photosensitive device, to mimic how the world is perceived by vision.

2) In computing, a display is described by the count of pixels across the width and height of the image. Conventional television would be denoted 644 × 483, which indicates 483 picture lines. But any display system involves some scanning overhead, so the total number of lines in the raster of conventional video is necessarily greater than 483.

3) Video scanning systems have traditionally been denoted by their total number of lines including sync and blanking overhead, the frame rate in hertz, and an indication of interlace (2∶1) or progressive (1∶1) scan. 525/59.94/2∶1 scanning is used in North America and Japan, with an analog bandwidth for studio video of about 5.5 MHz. 625/50/2∶1 scanning is used in Europe and Asia, with an analog bandwidth for studio video of about 6.5 MHz. 1125/60/2∶1 scanning is in use for high-definition television (HDTV), with an analog bandwidth of about 30 MHz.

4) A video system conveys image data in the form of a component that represents brightness, and two other components that represent color. It is important to convey the brightness component in such a way that noise introduced in transmission,

processing, and storage has a perceptually similar effect across the entire tone scale from black to white.

5) Although the shape of the retina is roughly a section of a sphere, it is topologically two-dimensional. In a camera, for practical reasons, we employ a flat image plane, instead of a spherical image surface. Image system theory concerns analyzing the continuous distribution of power that is incident on the image plane. A photographic camera has, in the image plane, film that is subject to chemical change when irradiated by light. The active ingredient of photographic film is contained in a thin layer of particles having carefully controlled size and shape, in a pattern with no coherent structure. If the particles are sufficiently dense, an image can be reproduced that has sufficient information for a human observer to get a strong sense of the original scene. The finer the particles and the more densely they are arranged in the film medium, the higher will be the capability of the film to record spatial detail.

6）人眼具有如下特性：出现在视网膜上的图像将保留几毫秒后才会消失。如果以每秒 50 幅的速率逐行扫描显示一组图像，人眼不会感觉是看到一幅一幅的离散图像。视频系统都是利用这一原理来产生运动画面的。

7）任何压缩方法都由两个算法构成。一个算法在信源压缩数据，另一个算法在信宿解压数据。在专业文献中，这两个算法分别称作“编码算法”和“解码算法”。

8）JPEG（联合图像专家组）用来对连续色调静止图像（如照片）进行压缩。JPEG 是由 ITU、ISO 和 IEC 共同支持的图像专家开发出来的。

9）宽高比是图像的宽度和高度之比。常规电视的宽高比是 4∶3。高清晰度电视采用的宽高比为 16∶9。摄像机常用的宽高比为 1.85∶1 或 2.35∶1。

10）MPEG 代表“移动图像专家组”，它是一组标准的统称。这些标准用来对数字压缩格式视听信息（如电影、录像、音乐）进行编码。与其他音、视频编码方式相比，MPEG 的主要优势是在相同质量下的文件要小得多。这是因为 MPEG 使用了非常复杂的压缩技术。

Reading Materials

Passage 1 Video on Demand (VOD)

Video on demand is sometimes compared to an electronic video rental store. The user (customer) selects any one of a large number of available videos and takes it home to view. Only with video on demand, the selection is made at home using the television set's remote control, and the video starts immediately. No trip to the store is needed. Needless to say, implementing video on demand is a wee bit more complicated than describing it.

Is video on demand really like renting a video, or is it more like picking a movie to watch from a 500-channel cable system? The answer has important technical implications. In particular, video rental users are used to the idea of being able to stop a video, make a quick trip to the kitchen or bathroom, and then resume from where the video stopped. Television viewers do not expect to put programs on pause.

If video on demand is going to compete successfully with video rental stores, it may be necessary to allow users to stop, start, and rewind videos at will. Giving users this ability virtually forces the video provider to transmit a separate copy to each one.

On the other hand, if video on demand is seen more as advanced television, then it may be sufficient to have the video provider start each popular video, say, every 10 minutes, and run these nonstop. A user wanting to see a popular video may have to wait up to 10 minutes for it to start. Although pause/resume is not possible here, a viewer returning to the living room after a short break can switch to another channel showing the same video but 10 minutes behind. Some material will be repeated, but nothing will be missed. This scheme is called near video on demand. It offers the potential for much lower cost, because the same feed from the video server can go to many users at once. The difference between video on demand and near video on demand is similar to the difference between driving your own car and taking the bus.

Watching movies on (near) demand is but one of a vast array of potential new services possible once wideband networking is available. The general model that many people use is illustrated in Figure 1. Here we see a high-bandwidth (national or international) wide area backbone network at the center of the system. Connected to it are thousands of local distribution networks, such as cable TV or telephone company

distribution systems. The local distribution systems reach into people's houses, where they terminate in set-top boxes, which are, in fact, powerful, specialized personal computers.

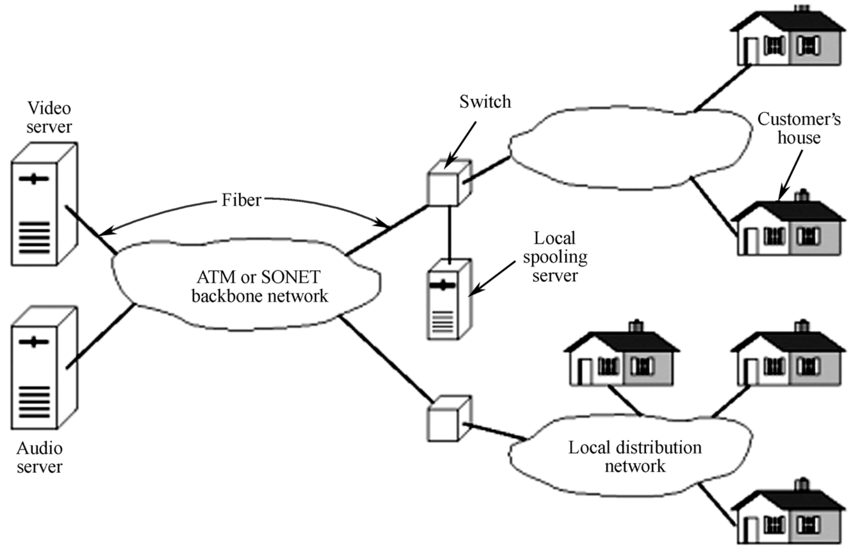


Figure 1 Overview of a video on demand system

Attached to the backbone by high-bandwidth optical fibers are numerous information providers. Some of these will offer pay-per-view video or pay-per-hear audio CDs. Others will offer specialized services, such as home shopping (letting viewers rotate a can of soup and zoom in on the list of ingredients or view a video clip on how to drive a gasoline-powered lawn mower). Sports, news, reruns of "I Love Lucy", WWW access, and innumerable other possibilities will no doubt quickly become available.

Also included in the system are local spooling servers that allow videos to be placed closer to the users (in advance), to save bandwidth during peak hours. How these pieces will fit together and who will own what are matters of vigorous debate within the industry.

Questions:

1) What does ***VOD*** stand for? And what's its meaning?

2) What're the technical implications when VOD is compared to an electronic video rental store?

3) How is a VOD system constructed?

4) What is near video on demand? And how is it different from the functions of VOD?

5) What's the backbone net work in a VOD system?

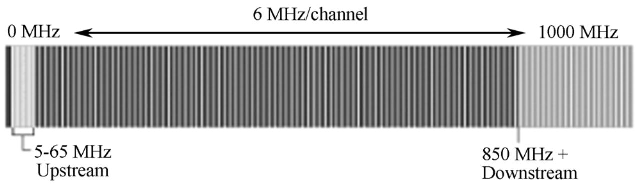
Passage 2　Cable Modems

For millions of people, television brings news, entertainment and educational programs into their homes. Many people get their TV signal from cable television (CATV) because cable TV provides a clearer picture and more channels.

Many people who have cable TV can now get a high-speed connection to the Internet from their cable provider. Cable modems compete with technologies like asymmetrical digital subscriber lines (ADSL). If you have ever wondered what the differences between DSL and cable modems are, or if you have ever wondered how a computer network can share a cable with dozens of television channels, then read on. This article will show you how a cable modem works and how 100 cable television channels and any Web site out there can flow over a single coaxial cable into your home.

Extra Space

You might think that a television channel would take up quite a bit of electrical "space" or bandwidth, on a cable. In reality, each television signal is given a 6-MHz channel on the cable. The coaxial cable used to carry cable television can carry hundreds of megahertz of signals—all the channels you could want to watch and more. In a cable TV system, signals from the various channels are each given a 6-MHz slice of the cable's available bandwidth and then sent down the cable to your house. In some systems(Figure 1), coaxial cable is the only medium used for distributing signals. In other systems, fiber-optic cable goes from the cable company to different neighborhoods or areas. Then the fiber is terminated and the signals move onto coaxial cable for distribution to individual houses.



Figuer 1　The Coaxial cable bandwidth

When a cable company offers Internet access over the cable, Internet information

can use the same cables because the cable modem system puts downstream data—data sent from the Internet to an individual computer—into a 6-MHz channel. On the cable, the data looks just like a TV channel. So Internet downstream data takes up the same amount of cable space as any single channel of programming. Upstream data -information sent from an individual back to the Internet—requires even less of the cable's bandwidth, just 2 MHz, since the assumption is that most people download far more information than they upload.

Putting both upstream and downstream data on the cable television system requires two types of equipment: a cable modem on the customer end and a cable modem termination system (CMTS) at the cable provider's end. Between these two types of equipment, all the computer networking, security and management of Internet access over cable television is put into place.

Inside the Cable Modem

Cable modems can be either internal or external to the computer. In some cases, the cable modem can be part of a set-top cable box, requiring that only a keyboard and mouse be added for Internet access. In fact, if your cable system has upgraded to digital cable, the new set-top box the cable company provides will be capable of connecting to the Internet, whether or not you receive Internet access through your CATV connection. Regardless of their outward appearance, all cable modems contain certain key components(Figure 2):

- A tuner
- A demodulator
- A modulator
- A media access control (MAC) device
- A microprocessor

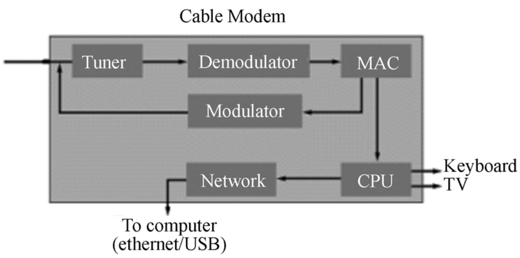


Figure 2 Cable modem components

Tuner

The tuner connects to the cable outlet, sometimes with the addition of a splitter that separates the Internet data channel from normal CATV programming. Since the Internet data comes through an otherwise unused cable channel, the tuner simply receives the modulated digital signal and passes it to the demodulator. In some cases, the tuner will contain a diplexer, which allows the tuner to make use of one set of frequencies (generally between 42 and 850 MHz) for downstream traffic, and another set of frequencies (between 5 and 42 MHz) for the upstream data. Other systems, most often those with more limited capacity for channels, will use the cable modem tuner for downstream data and a dial-up telephone modem for upstream traffic. In either case, after the tuner receives a signal, it is passed to the demodulator.

Demodulator

The most common demodulators have four functions. A quadrature amplitude modulation (QAM) demodulator takes a radio-frequency signal that has had information encoded in it by varying both the amplitude and phase of the wave, and turns it into a simple signal that can be processed by the analog-to-digital (A/D) converter. The A/D converter takes the signal, which varies in voltage, and turns it into a series of digital 1s and 0s. An error correction module then checks the received information against a known standard, so that problems in transmission can be found and fixed. In most cases, the network frames, or groups of data, are in MPEG format, so an MPEG synchronizer is used to make sure the data groups stay in line and in order.

Modulator

In cable modems that use the cable system for upstream traffic, a modulator is used to convert the digital computer network data into radio-frequency signals for transmission. This component is sometimes called a burst modulator, because of the irregular nature of most traffic between a user and the Internet, and consists of three parts:

- A section to insert information used for error correction on the receiving end
- A QAM modulator
- A digital-to-analog (D/A) converter

Media Access Control (MAC)

The MAC sits between the upstream and downstream portions of the cable modem, and acts as the interface between the hardware and software portions of the various network protocols. All computer network devices have MACs, but in the case of a cable modem the tasks are more complex than those of a normal network interface card. For

this reason, in most cases, some of the MAC functions will be assigned to a central processing unit (CPU) - either the CPU in the cable modem or the CPU of the user's system.

Microprocessor

The microprocessor's job depends somewhat on whether the cable modem is designed to be part of a larger computer system or to provide Internet access with no additional computer support. In situations calling for an attached computer, the internal microprocessor still picks up much of the MAC function from the dedicated MAC module. In systems where the cable modem is the sole unit required for Internet access, the microprocessor picks up MAC slack and much more. In either case, Motorola's PowerPC processor is one of the common choices for system designers.

Cable Modem Termination System(CMTS)

At the cable provider's head-end, the CMTS provides many of the same functions provided by the DSLAM in a DSL system. The CMTS takes the traffic coming in from a group of customers on a single channel and routes it to an Internet service provider (ISP) for connection to the Internet. At the head-end, the cable providers will have, or lease space for a third-party ISP to have, servers for accounting and logging, Dynamic Host Configuration Protocol (DHCP) for assigning and administering the IP addresses of all the cable system's users, and control servers for a protocol called CableLabs Certified Cable Modems—formerly Data Over Cable Service Interface Specifications (DOCSIS), the major standard used by U. S. cable systems in providing Internet access to users.

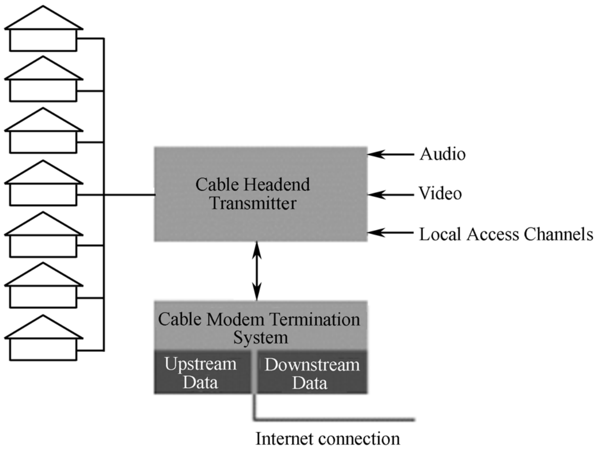


Figure 3 The cable modem termination system

The downstream information flows to all connected users, just like in an Ethernet network—it's up to the individual network connection to decide whether a particular block of data is intended for it or not. On the upstream side, information is sent from the user to the CMTS—other users don't see that data at all. The narrower upstream bandwidth is divided into slices of time, measured in milliseconds, in which users can transmit one "burst" at a time to the Internet. The division by time works well for the very short commands, queries and addresses that form the bulk of most users' traffic back to the Internet.

A CMTS will enable as many as 1000 users to connect to the Internet through a single 6-MHz channel. Since a single channel is capable of 30 to 40 megabits per second (Mbps) of total throughput, this means that users may see far better performance than is available with standard dial-up modems. The single channel aspect, though, can also lead to one of the issues some users experience with cable modems.

If you are one of the first users to connect to the Internet through a particular cable channel, then you may have nearly the entire bandwidth of the channel available for your use. As new users, especially heavy-access users, are connected to the channel, you will have to share that bandwidth, and may see your performance degrade as a result. It is possible that, in times of heavy usage with many connected users, performance will be far below the theoretical maximums. The good news is that this particular performance issue can be resolved by the cable company adding a new channel and splitting the base of users.

Another benefit of the cable modem for Internet access is that, unlike ADSL, its performance doesn't depend on distance from the central cable office. A digital CATV system is designed to provide digital signals at a particular quality to customer households. On the upstream side, the burst modulator in cable modems is programmed with the distance from the head-end, and provides the proper signal strength for accurate transmission.

Questions:

1) Why does people prefer a cable television?
2) Do you know the bandwidth for up-stream data in a cable modem system?
3) What components does a typical cable modem contain?
4) What functions does the demodulator usually perform in a cable modem?
5) What benefits does people get when using a cable modem for Internet access.

Passage 3 HDTV

HDTV has been getting media attention for several years now, and if you go to an electronics store you can see a fairly good selection of HDTV sets today. If you have ever looked at one of these sets, you know that the image they display is sharper and wider —it is more like a movie screen than it is a TV set!

HDTV has lifelike pictures and digital sound. The higher resolution produces clarity like you have never seen from a picture tube. Films retain their original width, enhancing your home theater experience. Imagine seeing more of a football field or a scenic panorama!

In analog TV, a 6 MHz analog signal carries intensity and color information for each scan line of the picture. An analog TV signal in the U. S. has 525 scan lines for the image, and each image is refreshed every 30th of a second (half of the scan lines are painted every sixtieth of a second in what is called an interlaced display). The horizontal resolution is something like 500 dots for a color set. This level of resolution was amazing 50 years ago, but today it is rather passe. The lowest resolution computer monitor that anyone uses today has 640×480 pixels, and most people use a resolution like 800×600 or 1024×768. We have grown comfortable with the great clarity and solidity of a computer display, and analog TV technology pales by comparison. Many of the new satellite systems, as well as DVDs, use a digital encoding scheme that provides a much clearer picture. In these systems, the digital information is converted to the analog format to display it on your analog TV. The image looks great compared to a VHS tape, but it would be twice as good if the conversion to analog didn't happen. There is now a big push underway to convert all TV sets from analog to digital, so that digital signals drive your TV set directly. When you read and hear people talking about digital television (DTV), what they are talking about is the transmission of pure digital television signals, along with the reception and display of those signals on a digital TV set. The digital signals might be broadcast over the air or transmitted by a cable or satellite system to your home. In your home, a decoder receives the signal and uses it, in digital form, to directly drive your digital TV set. There is a class of digital television that is getting a lot of press right now. It is called high-definition television, or HDTV. HDTV is high-resolution digital television (DTV) combined with Dolby Digital

surround sound (AC-3). HDTV is the highest DTV resolution in the new set of standards. This combination creates a stunning image with stunning sound. HDTV requires new production and transmission equipment at the HDTV stations, as well as new equipment for reception by the consumer. The higher resolution picture is the main selling point for HDTV. Imagine 720 or 1080 lines of resolution compared to the 525 lines people are used to in the United States (or the 625 lines in Europe) —it's a huge difference! Of the 18 DTV formats, six are HDTV formats, five of which are based on progressive scanning and one on interlaced scanning. Of the remaining formats, eight are SDTV (four wide-screen formats with 16 : 9 aspect ratios, and four conventional formats with 4 : 3 aspect ratios), and the remaining four are video graphics array (VGA) formats. Stations are free to choose which formats to broadcast. The formats used in HDTV are:

- 720p - 1280×720 pixels progressive
- 1080i - 1920×1080 pixels interlaced
- 1080p - 1920×1080 pixels progressive

"Interlaced" or "progressive" refers to the scanning system. In an interlaced format, the screen shows every odd line at one scan of the screen, and then follows that up with the even lines in a second scan. Since there are 30 frames shown per second, the screen shows one half of the frame every sixtieth of a second. For smaller screens, this is less noticeable. As screens get larger, the problem with interlacing is flicker. Progressive scanning shows the whole picture, every line in one showing, every sixtieth of a second. This provides for a much smoother picture, but uses slightly more bandwidth.

MPEG-2 & MPEG-4

Broadcasters are having to squeeze the increased picture detail and higher quality surround sound into the same 6-megahertz (MHz) bandwidth used by analog television. Compression software, very similar to what is used in personal computing, allows this to happen. Digital TV relies on a compression and encoding scheme known as MPEG-2 to fit its stunning images into a reasonable amount of bandwidth. In each image, the MPEG-2 software records just enough of the picture without making it look like something is missing. In subsequent frames, the software only records changes to the image and leaves the rest of the image as-is from the previous frame. MPEG-2 reduces the amount of data by about 55 to 1. MPEG-2 already is the industry standard for DVD videos and some of the satellite TV broadcast systems. Compression reduces image quality from what is seen by the digital camera at the studio. However, MPEG-2 is very

good at throwing away image detail that the human eye ignores anyway. The quality of the image is very good, and significantly better than traditional analog TV. The use of MPEG-2 permits an HDTV receiver to interact with computer multimedia applications directly. For example, an HDTV show could be recorded on a multimedia computer, and CD-ROM applications could be played on HDTV systems. A digital TV decodes the MPEG-2 signal and displays it just as a computer monitor does, giving it high resolution and stability. The MPEG-2 technology, however, still takes up a significant amount of bandwidth, and so satellite broadcasters have begun to embrace the newer technology of MPEG-4. MPEG-4 compression technology is able to squeeze twice as much HD video into the same amount of bandwidth as MPEG-2. Others will soon likely follow the satellite broadcaster's leads and slowly begin to embrace the newer MPEG-4 technology, making it the new standard.

HDTV Stations

There are HDTV stations "on the air" in many large cities. The first HDTV station was WRAL-HD in Raleigh, NC. The Federal Communications Commission (FCC) has mandated that all stations be capable of broadcasting HDTV by 2006. The timeline of HDTV coverage gives you an idea of what will be available in your area, and when.

The FCC mandate affects broadcasters, cable companies and consumers in significant ways:

- Consumers have to buy new equipment, either a set-top box (to convert digital signals to analog signals) or a whole new TV set.
- Broadcasters have to spend a considerable amount of money to switch to HDTV. They have to buy new cameras, new titling and editing equipment, new tape machines, new rigs for their news vans—it's a big investment.
- Cable operators have to convert all of their equipment and all of their set-top boxes.
- Communities need to agree to have new towers built for broadcast channels.

The station decides which DTV format it will transmit. For example, cable operators may push for 720p so they can fit more HDTV channels onto the cable. A clear pattern has yet to emerge in the industry.

How is HDTV Different?

The usual National Television Standards Committee (NTSC) analog TV screen in the U. S. has 525 scan lines, with 480 actually visible. The usual TV has an effective picture resolution of about 210,000 pixels. In the highest resolution digital TV formats, each picture contains about 2 million pixels. This means about 10 times more picture

detail on the HDTV screen!

The typical TV show uses 35-mm film (or is recorded direct-to-video using NTSC equipment). In the case of film, the broadcaster converts it to an analog TV signal for broadcasting. Standard 35-mm film has an aspect ratio of 1.37 : 1, meaning it is 1.37 times as wide is it high. A conventional TV screen has a 4 : 3 (1.33 : 1) aspect ratio, so the conversion is easy.

To deal with HDTV's new standards, broadcasters will need to get all new equipment, such as cameras, remote broadcast units, control rooms, cables, and sound equipment. This is because digital TV has:

- Wider images
- Much more detailed pictures
- 5.1 channel CD-quality Dolby Digital (AC-3) surround sound
- The ability to send data directly to a screen or to a PC as a download (The actual HDTV transmission is based on a 19.3-Mbps digital data stream.)

The aspect ratio (width to height) of digital TV is 16 : 9 (1.78 : 1), which is closer to the ratios used in theatrical movies, typically 1.85 : 1 or 2.35 : 1. Currently broadcasters must either pan and scan the image (crop the full picture of the film down to 4 : 3, eliminating part of every scene in the process) or letterbox it (present the full picture only on the middle part of the screen, with black bars above and below it). With a 16 : 9 screen, panning and scanning a theatrical movie doesn't remove so much from the original picture and letterboxing doesn't block out so much of your screen.

What Do I Need to Do?

When local over—the—air television stations moving to digital, you need to check your television sets. If your TV receives signals via cable, satellite or the internet, you will not be affected by the transition. You will continue to receive your existing television services. You do not need to buy additional equipment or subscribe to additional TV services.

If your TV receives signals over—the—air using an outdoor antenna or "Rabbit Ears", you may be affected by the transition to over—the—air digital television. If your television is equipped with a digital tuner, you will be able to continue viewing local stations that have switched to digital. The majority of television antennas that are currently used for watching analog signals will continue to work with digital signals. If your television does not have a built—in digital tuner, you have three options:

- Install a digital—to—analog converter box. A digital—to—analog converter box is a unit that sits on top of or near your television and that picks up over—the—

air digital signals to convert them for display on a standard analog television. However, the HDTV shows you see will look no better than DVD on your analog TV—you will get none of the resolution and format benefits of a real HDTV set.

- Buy a Digital Television. Today's HDTV sets come in several forms. Be sure any television receiver you purchase has input jacks that match the connectors on the VCR, cable box, DVD player and video game console you currently own. For many years, you will have to straddle the digital/analog fence, using, for example, an analog VCR on your digital TV. At the moment, there are no "standards" for what connections will appear on the back of an HDTV set. Therefore you should look for composite, S-video and component video as a minimum set of analog jacks so you can use your existing analog equipment with the new set.

  Many early purchasers will have to "go back" to a traditional outside UHF television antenna to receive the over-the-air (OTA) HDTV signal. The HDTV transmission system is an eight-level vestigial sideband (VSB) technique that uses UHF channels. Your antenna rotor setting for reception of HDTV signals will be easy to adjust. You either have a picture or you do not—there can not be a snowy image with digital technology. There also will not be any "fringe area" reception.

- Subscribe to Cable, Satellite, or an Internet—protocol (IP) Television Service.

Questions:

1) What does ***HDTV*** stand for? And what's its meaning?

2) Do you know the formats used in HDTV?

3) Why MPEG-2 is used in HDTV systems?

4) What kinds of new equipments should a broadcaster buy for HDTV's new standards? Why?

Unit 9

Embedded Applications

Lesson 25 Choosing the Right Core

Lesson 26 Design Languages for Embedded Systems

Lesson 27 Choosing a Real-Time Operating System

Passage 1 Personal Digital Assistant (PDA)

Passage 2 ARM

Passage 3 Embedded OS

Lesson 25 Choosing the Right Core

Generally speaking, there are seven very widely used 32-bit cores on the market at the moment: Motorola 680x0, Intel x86, PowerPC[1], MIPS[2], SuperH[3], and ARM[4]. Numerous less popular or proprietary architectures also exist, of course; many of these are associated with specific applications such as laser printers or DVD players.

Usage of the 680x0 cores appears to be in decline, and it is perhaps actually close to the end of its life; the principal consumer at this time is in PalmOS devices. These PDAs are now migrating towards ARM, and even Motorola has introduced an ARM-cored processor as its new flagship PDA part.

Architectures based around the high-end x86 family have some immediate advantages:

- You can use almost any PC-compatible operating system, and free software development tools.
- Installing operating systems is simple; in most cases there are automated installers that will probe your hardware combination and automatically install appropriate kernels, drivers etc. Compare this to the norm with embedded systems, where you will need to look at the board, work out the hardware configuration yourself, and sysgen [5] the kernel and driver set on external hardware, probably using a cross-compiler.
- It is simple to interface literally thousands of peripheral components for almost any imaginable function. Because these components are produced for the consumer market, with its enormous volumes and bloodthirsty price competition, peripheral components are cheap and fairly easy to acquire.
- Driver support exists (within the framework of most off-the-shelf operating systems) for almost any piece of hardware you could want to attach to your system.
- Highly integrated mainboards are available with many possible combinations of peripherals, in a wide variety of form factors.
- Migrating to a slightly different hardware platform due to shortages of support parts or evolving customer needs is relatively simple; in many cases, it simply involves recompiling and reinstalling the operating system and preparing a new

master disk image for duplication.

The obvious virtues of these parts having been extolled, some of the downsides must be pointed out:

- x86 parts are very expensive, in production quantities, compared to RISC alternatives of comparable performance. This may affect your ability to commercialize your device.
- There are relatively few x86 variants that are true "system on chip" devices, so you are likely to need quite a bit of external hardware in addition to the microprocessor itself. Often, in order to obtain one specific function, you will need to add a complex multifunction part because the single function you want isn't available as a discrete component. Again, this brings up your system complexity and total bill-of-materials cost.
- x86 has significant power consumption, heat and size disadvantages.
- Modern x86 parts and their support chips are very high-speed devices in dense packages. It is virtually impossible to hand-prototype your own design based around these parts; unless you want to spend many thousands of dollars on equipment, at the very least you will have to contract out some assembly work.
- PC peripheral ICs often have very short production lifespans; twelve to eighteen months is not uncommon, so ongoing sourcing may be an issue.
- Code to cold-boot a "bare" PC platform is usually very complicated, because you have to replace numerous layers—motherboard BIOS, expansion card BIOS, and various OS layers. The CPU architecture is also complex.
- It bears pointing out that JTAG[6]-based or other hardware debugging systems aren't usually available on commercial single-board x86 computers.

I recommend x86 as the platform of choice if you are either building just a few of your appliance, or if you are prototyping something and want to pull together a lot of miscellaneous hardware features without spending a lot of time debugging the hardware design. It's also a good choice for an initial production run that you can ship to early adopters while you are developing a cheaper second-round customized hardware design. There are other special situations where you might find x86 to be a good choice, but these are the major ones.

Moving onto the RISC platforms, MIPS, SuperH and PowerPC are good candidates for many applications, and in particular the SuperH family is large and contains a wide variety of useful devices, though MIPS seems to be a more widely licensed core in third-party ASICs and ASSPs. PowerPC seems to be found mainly in applications requiring

very high performance. In evaluating all of these parts for various projects, I have found them to be fairly difficult to develop with on a shoestring budget; evaluation hardware is usually costly, and most variants of these parts are not readily available to buyers who are unable to demonstrate a need for large quantities. However, all of these cores are likely to remain available and well-supported for the foreseeable future, so they are all viable choices as long as you can obtain development systems and parts.

At least in the case of SuperH and MIPS, your cheapest path to a prototype based on these parts is generally to repurpose some existing piece of hardware such as a PDA; for PowerPC, I would suggest buying a commercial single-board industrial control computer based around the chip of interest. Be warned that this is likely to be expensive; PowerPC boards don't have the same kind of mass-market pricing as x86-compatible boards and you can expect to pay between two and three times as much for a PowerPC SBC as for a comparable x86-based board.

Bearing the above discussion in mind, unless some of the Intel arguments apply to your case my primary recommendation for a 32-bit embedded platform is ARM. This architecture has many important advantages (some of these are also applicable to the other RISC platforms mentioned above, of course):

- It is a mature, well-understood architecture with a solid engineering history and many refinements. The large number of current licensees and now-shipping parts makes ARM a very safe bet for future availability.
- The cores are small and have excellent power consumption vs. performance characteristics.
- Many features-coprocessors, external bus widths, memory management unit, cache size, etc. are tunable by the chip designer, meaning that a core variant can be found to meet almost any performance/size/power requirement.
- There are a huge number of attractively priced standard, custom and semi-custom parts on the market with a wide variety of integrated peripherals.
- Since ARM provides reference designs for many different peripherals as well as the core itself, there are often similarities in peripheral control on different ARM implementations, even from different vendors. To take a trivial example, code to send data out of a serial port can usually be ported from one ARM variant to another with little effort.
- Partly because of the above factors, there is a huge amount of freely available intellectual property already extant for this core.

The cliché often used is that "ARM is the 32-bit 8051," meaning that it is the universal

32-bit microcontroller core known to everybody and used everywhere. This is barely an exaggeration; ARM is to the embedded world what x86 is to the desktop PC world.

New Words

flagship [ˈflæɡʃip] *n.* 旗舰
probe [prəub] *vt.* 探查,查明
norm [nɔːm] *n.* 标准,规范
framework [ˈfreimwəːk] *n.* 框架,结构
extol [ikˈstəul] *v.* 赞美
downside [ˈdaunˌsaid] *n.* 不利方面
multifunction [ˌmʌltiˈfʌŋkʃn] *n.* 多功能
lifespan [ˈlaifspæn] *n.* 预期使用期限,平均生命期,存在时间
uncommon [ʌnˈkɔmən] *adj.* 不凡的,罕有的,难得的
ongoing [ˈɔnɡəuiŋ] *adj.* 正在进行的
shoestring [ˈʃuːstriŋ] *n.* 鞋带,小额资本
viable [ˈvaiəbl] *adj.* 能生存的,可行的
refinement [riˈfainmənt] *n.* 改进
licensee [ˌlaisənˈsiː] *n.* 获许可的人,技术引进方
variant [ˈvɛəriənt] *n.* 变体
extant [ekˈstænt] *adj.* 现存的,未毁的
cliché [ˈkliːʃei] *n.* 空话,套话,废话

Phrases & Expressions

off the shelf 现货供应的
work out 设计出,做出
bring up 提出,引出
contract out 包出

Technical Terms

cross-compiler *n.* 交叉编译器
bill of materials 材料单
PDA *abbr.* Personal Digital Assistant 个人数字助理

RISC [risk] *abbr*. Reduced Instruction Set Computer 精简指令集计算机
JTAG ['dʒeitæg] *abbr*. Joint Test Action Group 联合测试工作组
ASIC ['eitsik] *abbr*. Application Specific Integrated Circuit 专用集成电路
ASSP *abbr*. Application Specific Standard Product 专用标准器件
IP *abbr*. Intellectual Property 知识产权

Notes

1. PowerPC 是 1991 年由 Apple、IBM、Motorola 联合推出的 RISC 微处理器架构。
2. MIPS（Microprocessor without Interlocked Pipeline Stages）是 MIPS 公司推出的 RISC 微处理器架构。
3. SuperH（或 SH）是 20 世纪 90 年代初由日立（Hitachi）公司推出的 32 位 RISC 微处理器架构。
4. ARM（Advanced RISC Machine）是由 ARM 公司推出的 RISC 微处理器架构。
5. sysgen 这个词是由 system 和 generation 合成的，表示"系统生成"的意思。
6. JTAG 即 IEEE 1149.1 标准，此标准使用"边界扫描"技术测试印制电路板。

Lesson 26　Design Languages for Embedded Systems

An embedded system is a computer masquerading as a non-computer that must perform a small set of tasks cheaply and efficiently. A typical system might need to do communication, signal processing, and user-interface tasks. Because the tasks must solve diverse problems, a general-purpose language to solve all embedded applications would be difficult to write, analyze, and compile. Instead, developers have introduced a variety of languages, each best suited to a particular problem domain.

Hardware Languages

Verilog and VHDL are the most popular languages for hardware description and modeling. Each hardware description language (HDL) models systems with discrete-event semantics that ignore idle portions of the design for efficient simulation. Both describe systems with structural hierarchy: a system consists of blocks that contain instances of primitives, other blocks, or concurrent processes. In addition, each HDL explicitly lists connections.

Verilog provides more primitives geared specifically toward hardware simulation. On the other hand, VHDL's primitives are assignments such as a=b+c or procedural code. Verilog adds transistor and logic-gate primitives, and lets you define new primitives with truth tables.

Both languages allow concurrent processes to be described procedurally. Such processes sleep until awakened by an event that causes them to run, read and write variables, and suspend. Processes may wait for a period of time (for example, #10 in Verilog, wait for 10ns in VHDL), a value change (@(a or b), wait on a, b), or an event (@(posedge clk), wait on clk until clk='1').

VHDL communication is more disciplined and flexible. Verilog communicates through wires or regs: shared memory locations that can cause race conditions [1]. VHDL's signals behave like wires, but the resolution function may be user-defined. VHDL's variables are local to a single process unless declared as shared variables.

Verilog's type system models hardware with four-valued bit vectors and arrays for modeling memory. VHDL does not include four-valued vectors, but its type system lets a user add them. Furthermore, you can define composite types such as C structs.

Overall, Verilog is the leaner language, more directly geared toward simulating digital integrated circuits. VHDL is a much larger, more verbose language capable of handing a wider class of simulation and modeling tasks.

Software Languages

Software languages describe sequences of instructions for a processor to execute. As such, most languages list imperative instructions, executed in order that communicate through memory-an array of storage locations that hold their values until changed.

Each machine instruction typically does little more than, for example, add two numbers, so high-level languages aim to specify many instructions concisely and intuitively. Arithmetic expressions are typical—coding an expression such as ax^2+bx+c in machine code is straightforward, tedious, and best done by a compiler. The C language provides such expressions, control-flow constructs such as loops and conditionals, and recursive functions. In addition, the C++ language provides classes as a way to build new data types, templates for polymorphic code, exceptions for error handling, and a standard library of common data structures. Java is a still higher-level language that provides automatic garbage collection, threads, and monitors to synchronize the threads.

Assembly Languages

An assembly language program is a list of processor instructions written in a symbolic, human-readable form. Each instruction consists of an operation, such as addition, along with some operands. For example, add r5, r2, r4 might add the contents of registers r2 and r4 and then write the result to r5. An assembly language executes these types of arithmetic instructions in order, but branch instructions can perform conditionals and loops by changing the processor's program counter — the address of the instruction being executed.

A processor's assembly language is defined by its opcodes, addressing modes, registers, and memories. The opcode distinguishes, for example, addition from a conditional branch, and an addressing mode defines how and where data is gathered and stored (such as, from a register or from a particular memory location). Registers can be thought of as small, fast, and easy-to-access pieces of memory.

The C Language

A C program contains functions built from arithmetic expressions structured with loops and conditionals. Instructions in a C program run sequentially, but control flow constructs such as loops of conditionals can affect the order in which instructions execute. When control reaches a function call in an expression, the program passes control to the called function. This function runs until it produces a result, and control returns to continue evaluating the expression that called the function.

C derives its types from those a processor manipulates directly: signed and unsigned integers ranging from bytes to words, floating point numbers, and pointers. These can be further aggregated into arrays and structures—groups of named fields.

C programs use three types of memory. The program allocates space for global data upon program compilation. The stack stores automatic variables the program allocates and releases the variables when the program calls the function and returns a value. The heap supplies arbitrarily sized regions of memory that the program can deallocate in any order. Although the C language is an ISO standard, many people consult the book by Kernighan and Ritchie, since Ritchie designed the language.

C++

C++ extends C with structuring mechanisms for big programs. These mechanisms include user-defined data types, a way to reuse code with different types, names to group objects and avoid accidental name collisions when program pieces are assembled, and exceptions to handle errors. The C++ standard library includes a collection of efficient polymorphic data types such as arrays, trees, and strings for which the

compiler generates custom implementations.

A class defines a new data type by specifying its representation and the operations that may access and modify it. Classes may be defined by inheritance, which extends and modifies existing classes. For example, a rectangle class might add length and width fields and an area method to a shape class.

A template is a function or class that can work with multiple types. The compiler generates custom code for each different use of the template. For example, a program could use the same min template for both integers and floating-point numbers.

Java

Sun's Java language resembles, but is incompatible with, C++ . Like C++ , Java is object-oriented, providing classes and inheritance. It is a higher-level language than C++ , since it uses object references, arrays, and strings instead of pointers. Java's automatic garbage collection frees the programmer from memory management.

Java provides concurrent threads. Creating a thread involves extending the Thread class, creating instances of these objects, and calling their start methods to start a new thread of control that executes the objects' run methods.

Synchronizing a method or block uses a per-object lock to resolve contention when two or more threads attempt to access the same object simultaneously. A thread that attempts to gain a lock owned by another thread will block until the lock is released—you can use this feature grant a thread exclusive access to a particular object.

RTOS

Many embedded systems use a real-time operating system (RTOS) to simulate concurrency on a single processor. An RTOS manages multiple running processes, each written in a sequential language such as C. The processes perform the system's computations and the RTOS schedules them. The scheduling attempts to meet deadlines by deciding which processes run at what times and in what order.

Most real-time operating systems use fixed-priority preemptive scheduling in which each process is given a particular priority (a small integer) when the system is designed. At any time, the RTOS runs the highest-priority running process, which is expected to run for a short period of time before suspending itself to wait for more data. Priorities are usually assigned using rate-monotonic analysis, which assigns higher priorities to processes that must meet deadlines that are more frequent.

New Words

masquerade [ˌmɑːskəˈreid] *n.* 伪装，乔装
suited [sjuːtid] *adj.* 适合的，匹配的
semantics [siˈmæntiks] *n.* 语义学
hierarchy [ˈhaiərɑːki] *n.* 层次，层级
explicitly [ikˈsplisitli] *adv.* 明确地
gear [ɡiə] *v.* 调整
disciplined [ˈdisiplind] *adj.* 受过训练的，遵守纪律的
verbose [vəːˈbəus] *adj.* 详细的，冗长的
imperative [imˈperətiv] *adj.* 强制的，必须的
concisely [kənˈsaisli] *adv.* 简明地
intuitively [inˈtjuitivli] *adv.* 直觉地，直观地
expression [ikˈspreʃn] *n.* 表达式
straightforward [ˌstreitˈfɔːwəd] *adj.* 易懂的，直接的
tedious [ˈtiːdiəs] *adj.* 乏味的，冗长的
construct [kənˈstrʌkt] *n.* 结构，构想，建造物
loop [luːp] *n.* 循环
conditional [kənˈdiʃnəl] *adj.* 条件的
template [ˈtemplit] *n.* 模板
polymorphic [ˌpɔliˈmɔːfik] *adj.* 多态的
symbolic [simˈbɔlik] *adj.* 符号的
integer [ˈintidʒə] *n.* 整数
aggregate [ˈæɡriɡeit] *v.* 聚集，合计
allocate [ˈæləukeit] *vt.* 分配
stack [stæk] *n.* 堆栈
field [fiːld] *n.* 字段
array [əˈrei] *n.* 数组，阵列
structure [ˈstrʌktʃə] *n.* 结构体
monotonic [mɔnəˈtɔnik] *adj.* 单调的
extend [ikˈstend] *v.* 扩展，扩充
mechanism [ˈmekənizəm] *n.* 机制
accidental [ˌæksiˈdentl] *adj.* 意外的
inheritance [inˈheritəns] *n.* 继承

collision [kəˈliʒn] *n.* 碰撞，冲突
schedule [ˈskedʒul] *v.* 调度
release [riˈliːs] *v.* 释放，发布

Phrases & Expressions

consist of… 由……组成
(be) capable of … 具备……的能力
be thought of as … 被认为……

Technical Terms

instance [ˈinstəns] *n.* 实例
primitive [ˈprimitiv] *n.* 原语，图元，基本元件
assignment [əˈsainmənt] *n.* 赋值
procedural [prəˈsiːdʒərəl] *adj.* 程序上的
user-defined *adj.* 自定义的
compiler [kəmˈpailə] *n.* 编译器
operand [ˈɔpəˌrænd] *n.* 操作数
function [ˈfʌŋkʃn] *n.* 函数，子程序
pointer [ˈpɔintə] *n.* 指针
deadline [ˈdedlain] *n.* 截止时间，完成期限
preemptive [priːˈemptiv] *adj.* 先占的
class [klɑːs] *n.* 类
string [striŋ] *n.* 字符串
thread [θred] *n.* 线程
contention [kənˈtenʃn] *n.* 争用
concurrent process 并发进程
truth table 真值表
resolution function 判决函数
bit vector 位向量
branch instruction 分支指令
program counter 程序计数器
addressing mode 寻址模式
floating point numbers 浮点数

global data 全局数据
automatic variable 自动变量
program call 程序调用
error handling 错误处理
object-oriented 面向对象的
object reference 对象引用
memory management 内存管理
HLL *abbr*. High-Level Language 高级语言
ISO *abbr*. International Standard Organization 国际标准化组织
RTOS *abbr*. Real-Time Operating System 实时操作系统

Notes

1. 竞争状态(race condition)是指当两个实体(如两个进程)对同一资源进行竞争时，因系统没有判定执行顺序的机制而导致的结果不可预测状态。
2. Java 是 Sun 公司(Sun Microsystems)于 1995 年推出的一种面向对象的、跨平台解释性程序设计语言，其语法规则和C++ 类似。Java 已成为互联网上广泛使用的编程语言之一。

Lesson 27 Choosing a Real-Time Operating System

Engineers often use the term"real time"to describe computing problems for which a late answer is as bad as a wrong one. These problems are said to have deadlines, and embedded systems frequently operate under such constraints. For example, if the embedded software that controls your anti-lock brakes misses one of its deadlines you might find yourself in an accident. So it's extremely important that the designers of real-time embedded systems know everything they can about the behavior and performance of their hardware and software.

The designers of real-time systems spend a large portion of their time worrying about worst-case performance. They must constantly ask themselves questions like: what is the worst-case time between the human operator pressing the brake pedal and an interrupt signal arriving at the processor? What is the worst-case interrupt latency? And what is the worst-case time for the software to respond by triggering the braking

mechanism? Average or expected-case analysis simply will not suffice. Most of the commercial embedded operating systems available today are designed for possible inclusion in real-time systems. In the ideal case, their worst-case performance is well understood and documented.

To earn the distinctive title "Real-Time Operating System" (RTOS), an operating system should be deterministic and have guaranteed worst-case interrupt latency and context-switch times. Given these characteristics and the relative priorities of the tasks and interrupts in your system, it is possible to analyze the worst-case performance of the software using a technique such as rate monotonic analysis.

An operating system is said to be deterministic if the worst-case execution time of each of its system calls is calculable. An operating system vendor that takes the real-time behavior of its product seriously will usually publish a datasheet providing the minimum, average, and maximum number of clock cycles required by each system call. These numbers may be different for different processors, but it is reasonable to expect that if the algorithm is deterministic on one processor it will be so on any other. (The actual times may differ, however.)

Interrupt latency is the total length of time from an interrupt signal arriving at the processor to the start of the associated interrupt service routine (ISR). When an interrupt occurs, the processor must take several steps before executing the ISR. First, the processor must finish executing the current instruction. Next, the interrupt type must be recognized. This is done by the hardware and does not slow or suspend the running task. Finally, and only if interrupts are enabled, the CPU's context is saved and the ISR associated with the interrupt is started.

Of course, if interrupts are ever disabled (say within a system call), the worst-case interrupt latency increases by the maximum amount of time that they are turned off. Each operating system will internally disable interrupts in several places and for different lengths of time, so it is important that you know what your system's requirements are. One real-time project might require a guaranteed interrupt response time as short as 1 μs, while another may require only 100 μs.

The third real-time characteristic of an operating system is the amount of time required to perform a context switch. This is important because it represents overhead across your entire system. For example, imagine if the average execution time of any task before it blocks is 100 ms but that the context-switch time is also 100 ms. In that case, fully one half of the processor's time is spent within the context-switch routine!

Again, there is no magic number and the actual times are usually processor-

specific, since they are dependent on the number of registers that must be saved and where. Be sure to get these numbers from any operating system vendor you are thinking of using. That way, there won't be any last minute surprises.

Selection process

Considering the cost of engineering time these days, a few thousand dollars is a bargain for a commercial RTOS. A wide variety of operating systems are available to suit most projects and pocketbooks. Commercial operating systems form a continuum of functionality, performance, and price. Those at the lower end of the spectrum offer just a basic preemptive scheduler and a few other key system calls. These operating systems are usually inexpensive, come with source code that you can modify, and do not require payment of any royalties. Operating systems at the other end of the spectrum typically include a lot of functionality beyond the basic scheduler. These operating systems can be quite expensive, though, with startup costs ranging from \$10,000 to \$50,000 and royalties due on every copy shipped in ROM. However, this price often includes free technical support and training and a set of integrated development tools. Between these two extremes are the operating systems that have modest up-front costs and/or royalties, but do not include source code; technical support may cost extra. Most commercial operating systems fall in this category.

With such a variety of operating systems and features to choose from, it can be difficult to decide which is the best for your project. Try putting your processor, real-time performance, and budgetary requirements first. These are criteria that you cannot change, so you can use them to narrow the possible choices to a smaller set of products. Then contact all of the vendors of the remaining operating systems for more detailed technical information.

At this point, many people make their decision based on compatibility with their choice of cross-compiler, debugger, and other development tools. But it's really up to you to decide what additional features are most important for your project. No matter what you decide to buy, the basic kernel will be about the same. The differences will most likely be measured in processor support, minimum and maximum memory requirements, availability of add-on software modules (for example, networking protocol stacks and device drivers), and compatibility with third-party development tools.

The best reason to choose a commercial operating system is the advantage of using something that is better tested and, therefore, more reliable than a kernel you have

developed in house. So one of the most important things you should be looking for from your OS vendor is experience.

New Words

constraint [kən'streint] *n*. 约束条件
latency ['leitənsi] *n*. 反应时间
brake [breik] *n*. 刹车
pedal ['pedl] *n*. 踏板
suffice [sə'fais] *vi*. 足够
inclusion [in'klu:ʒn] *n*. 包括,蕴含
well-documented [ˌwel'dɔkjuməntid] *adj*. 备有证明文件的,有执照的
deterministic [diˌtə:mi'nistik] *adj*. 确定性的
context ['kɔntekst] *n*. 上下文,环境
calculable ['kælkjuləbl] *adj*. 可计算的,能预测的
bargain ['bɑ:gin] *n*. 交易,协议,廉价品
pocketbook ['pɔkitˌbuk] *n*. 笔记本,钱袋
continuum [kən'tinjuəm] *n*. 连续区间
spectrum ['spektrəm] *n*. 频谱,领域,范围
royalty ['rɔiəlti] *n* 特许权,专利权税,转让费
extreme [ik'stri:m] *n*. 极端,末端
upfront [ʌp'frʌnt] *adj*. 在前面的,提前支付的
budgetary ['bʌdʒitəri] *adj*. 预算的
criteria [krai'tiəriə] *n*. 标准
network ['netwə:k] *n*. 网络
inhouse *adj*. 内部的,自身的,固有的,自用的

Phrases & Expressions

a portion of 一部分
a variety of 各种各样的

Technical Terms

algorithm ['ælgəriðəm] *n*. 算法

disable [dis'eibl] *v.* 禁用，使无效
scheduler ['ʃedjuːlə] *n.* 调度程序
startup *n.* 启动
debugger [diː'bʌgə] *n.* 调试器
kernel ['kəːnl] *n.* 内核
add-on *n.* 附件
stack [stæk] *n.* 堆栈
protocol ['prəutəkɔl] *n.* 协议，规程
interrupt latency 中断响应时间
system call 系统调用
execution time 执行时间
source code 源代码，源程序
cross compiler 交叉编译器
device driver 设备驱动程序
RTOS *abbr.* Real-Time Operating System 实时操作系统
ISR *abbr.* Interrupt Service Routine 中断服务程序
IDE *abbr.* Integrated Development Environment 集成开发环境

Notes

1. 算法是指一种在有限步骤内解决问题的计算程序。
2. 任务的反应时间(latency)或迟滞时间(tardiness)是指其实际启动(或结束)的时刻和应该启动(或结束)的时刻之差。决定反应时间的主要因素包括处理器、总线、存储器和外设的时序特性，操作系统的调度特性，操作系统内核的抢占性，系统载荷和任务切换时间。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) An embedded system is any electronic system that uses a CPU chip, but that is not a ______(通用的)workstation, desktop ______ laptop computer. ______(这样的系统一般使用微处理器), or they may use ______(定制芯片)or both. They are used in automobiles, planes, trains, space vehicles, machine tools, cameras, consumer and office appliances, cell-phones, PDAs and other hand-helds ______ robots and toys. The uses are endless, and billions of microprocessors are shipped every year for a myriad of

applications. Although there are embedded versions of popular ______(操作系统), ______(低成本消费类产品)can use chips that cost less than a dollar and have ______(非常有限的指令存储器). ______(在此类情况下), the OS and application may be combined into one program.

In embedded systems, the software is permanently set into a read-only memory such as a ROM or flash memory chip, in contrast ______ a general-purpose computer that ______(将其程序载入随机存取存储器中)each time. Sometimes, single board and rack mounted general-purpose computers ______(被称做嵌入式计算机)if used to control a single printer, drill press or other such device.

(2) Multithreading means multitasking within ______(一个程序). It allows multiple streams of execution to ______(并行发生) within the same program, each stream processing a different ______(事务或者消息). In order ______(使一个多线程程序性能获得真正提高), it must be run ______ a multitasking or multiprocessing environment, which allows multiple operations ______ take place.

Certain types of applications lend themselves to multithreading. For example, in an order processing system, each order can be entered independently of the other orders. In an image editing program, a ______(运算密集的)filter can be performed on one image, while the user works on another. In a ______(对称的)multiprocessing (SMP) operating system, its multithreading allows multiple CPUs to be controlled ______ the same time. It is also used to create synchronized ______(音视频) applications.

Multithreading generally uses ______(重入代码), which cannot be modified when executing, ______ the same code can be shared by multiple programs.

2. Translate the following passages into Chinese or English.

1) Hardware means machinery and equipment (CPU, disks, tapes, modem, cables, etc.). In operation, a computer is both hardware and software. One is useless without the other. The hardware design specifies the commands it can follow, and the instructions tell it what to do.

2) Firmware is a category of memory chips that hold their content without electrical power and include ROM, PROM, EPROM and EEPROM technologies. Firmware becomes "hard software" when holding program code.

3) Software is "logic and language". Software deals with the details of an ever-changing business and must process transactions in a logical fashion. Languages are used to program the software. The "logic and language" involved in analysis and programming is generally far more complicated than specifying a storage and

transmission requirement.

4) Software package is an application program developed for sale to the general public. Packaged software is generally designed to appeal to a large audience of users, and although the programs may be tailored to a user's taste by setting various preferences, it is not as individualized as custom-designed and custom-programmed software.

5) Process is the unique execution of a particular piece of code by a specific user on the same computer. Thus, if Alice runs the Microsoft WORD at 8:30 and then again at 8:36, she has generated two processes. If Alice and Bob both run the WORD at 8:30 on the same computer, they have also generated two processes.

6)“实时系统”是这样一种计算机系统:它能够对输入信号做出快速响应来维持所需的处理速度。

7)“实时操作系统”是专为“实时计算机系统”设计的操作系统。

8)系统软件是用于控制计算机和开发、运行应用程序的程序。系统软件包括操作系统、TP 监控程序、网络操作系统和数据库管理器。

9)“硬件”意味着“存储和传输”。计算机拥有的内存和磁盘存储器越多,它就能做更多的工作。内存和磁盘向 CPU 传送数据和指令的速度越快,CPU 完成得也越快。对于硬件的需求取决于将创建的数据库大小、用户数量或需同时服务的应用程序数量。

10)“软件”意味着“计算机指令”。“程序”就是用以完成某项工作的一组指令。软件可分为“系统软件”、“应用软件”两大类。“系统软件”由操作系统、数据库管理系统等控制程序组成,而用户处理数据的任何程序(如电子表格软件、字处理软件等)都属于“应用软件”。人们常把“软件”误认为是“数据”,其实不然。软件告诉硬件该如何处理数据。软件是“被运行的”,而数据是“被处理的”。

Reading Materials

Passage 1 Personal Digital Assistant (PDA)

The main purpose of a personal digital assistant (PDA) is to act as an electronic organizer or day planner that is portable, easy to use and capable of sharing information with your PC. It's supposed to be an extension of the PC, not a replacement.

PDAs have definitely evolved over the years. Not only can they manage your personal information, such as contacts, appointments, and to-do lists, today's devices can also connect to the Internet, act as GPS devices, and run multimedia software. What's more, manufacturers have combined PDAs with cell phones, multimedia players and other electronic gadgetry(See Table 1).

Table 1 Types of PDAs

Types	Description	Features
Palm PDAs	Most Palm devices are made by palmOne	
Pocket PCs	Pocket PC is the generic name for Windows Mobile PDAs	• Pocket versions of Microsoft applications Three handwriting-recognition applications • A virtual writing area • Windows Media Player for multimedia content
Smartphones	A smartphone is either a cell phone with PDA capabilities or a traditional PDA with added cell phone capabilities	• A cellular service provider to handle phone service • Internet access through cellular data networks • Various combinations of cell phone and PDA features • A number of different operating systems

PDAs mostly evolved from stand—alone devices to a handful of applications in a smartphone, which can also perform countless other tasks. Some newer models have audio capabilities, enabling them to be used as mobile phones or portable media players. Wireless PDAs may also offer e-mail and Web browsing, and data are synchronized between the PDA and desktop computer via USB or wireless. Many PDAs employ touchscreen technology.

Microprocessors and Memory

Like standard desktop and laptop computers, PDAs are powered by microprocessors(See Figure 1). Unlike desktop PCs and laptops, PDAs use smaller, cheaper microprocessors. Although these microprocessors tend to be slower than their PC counterparts, they are adequate for the tasks that PDAs perform. The benefits of small size and price outweigh the cost of slow speeds.

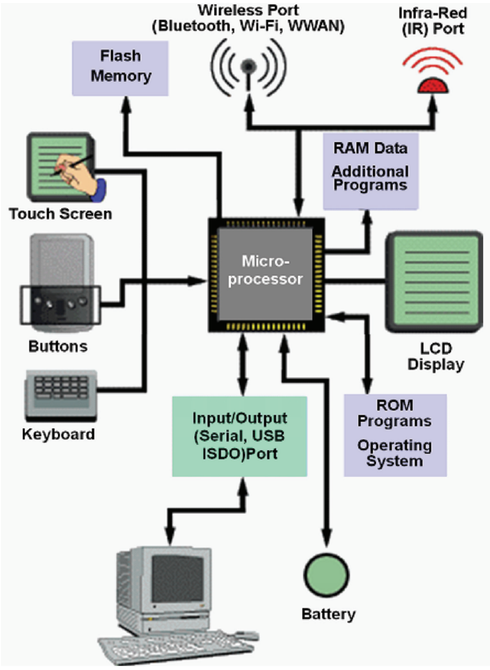


Figure 1 The parts that can make up a PDA

A PDA doesn't have a hard disk. It stores basic programs (address book, calendar, memo pad and operating system) in a ROM chip, which remains intact even when the machine shuts down. Your data and any programs you add later are stored in the device's RAM. Information in RAM is only available when the device is on. Due to their design, PDAs keep data in RAM safe because they continue to draw a small amount of power from the batteries even when you turn the device off.

Less powerful PDAs have lower amounts of RAM. However, many application programs take up significant memory space, so most models have more memory. Also, Pocket PC devices generally require more resources and have even more RAM. To provide additional memory, many PDAs accept removable flash media add-on cards.

These are handy for storing large files or multimedia content, such as digital photos.

Some newer PDAs, such as the Palm Tungsten E2, use flash memory instead of RAM. Flash memory is non-volatile, which means it preserves the data and applications it stores-even when all battery power is depleted.

Operating Systems

The operating system contains the pre-programmed instructions that tell the microprocessor what to do. The operating systems used by PDAs are not as complex as those used by PCs. They have fewer instructions, which require less memory.

PDAs and smartphones typically have one of two types of operating systems: Palm OS or Windows Mobile. However, RIM makes a specific OS for its BlackBerry devices, and the Symbian OS operates some smartphones.

Batteries

PDAs are powered by batteries. Some models use alkaline (AAA) batteries, while others use rechargeable batteries (lithium, nickel-cadmium or nickel-metal hydride). The battery life depends on what kind of PDA you have and how you use it. Here are some of the things that can drain batteries:

- Operating system (PocketPC requires more power by virtue of its increased memory requirements)
- More memory
- Wireless connections, such as Wi-Fi and Bluetooth
- Backlighting on the display

Battery life can vary from hours to months, depending upon the PDA model and its features. Most PDAs have power management systems in place to extend the battery life. Even if the batteries are so low that you can no longer turn the machine on (it will give you plenty of warning before this happens), there's usually enough power to keep the RAM refreshed.

If the batteries do run completely out of juice or if you remove them, most devices have an internal backup battery that provides short-term power (typically 30 minutes or less) until you install a replacement.

In addition to battery power, many PDAs come with AC adapters to run off household electric current. A car adapter is also generally available as an accessory.

LCD Display

PDAs use an LCD screen. Unlike the LCD screens for desktop or laptop computers, which are used solely as output devices, PDAs use their screens for output and input. The LCD screens of PDAs are smaller than laptop screens, but vary in size. Almost all PDAs now offer color displays.

PDA displays have the following features:

- Transflective TFT (thin-film transistor) LCD for indoor and outdoor use
- Different pixel resolutions with higher resolutions for better quality
- Color screen
- Backlighting for reading in low light

Input Methods

PDAs vary in how you input data and commands. Some devices use a stylus and touch screen exclusively in combination with a handwriting recognition program. Using a plastic stylus, you draw characters on the device's display or dedicated writing area. Software inside the PDA converts the characters to letters and numbers. On Palm devices, the software that recognizes these letters is called Graffiti. Graffiti requires that each letter be recorded in a certain way, and you must use a specialized alphabet. For example, to write the letter "A", you draw an upside-down V. The letter "F" looks like an inverted L. To help Graffiti make more accurate guesses, you must draw letters on one part of the screen and numbers in another part.

Pocket PC PDAs offer three handwriting-recognition applications: Transcriber, Letter Recognizer and Block Recognizer. Letter Recognizer and Block Recognizer are similar to Graffiti and require specialized alphabets. By contrast, Transcriber recognizes your "regular" handwriting, as long as you write legibly. It is similar to the handwriting recognition capabilities found on Tablet PCs.

If you can't get the hang of PDA handwriting, you can use a miniature onscreen keyboard. It looks just like a regular keyboard, except you tap on the letters with the stylus. In addition, many devices now include a small (and usually cramped) QWERTY keyboard. Some of these require you to use your thumbs to type. And you can use a full-size keyboard by connecting it to the PDA via Bluetooth or a USB port. Each model also has a few buttons and navigation dials to bring up applications and scroll through files.

Touch screen

Many of the original PDAs, such as the Apple Newton and Palm Pilot, featured a touchscreen for user interaction, having only a few buttons—usually reserved for shortcuts to often-used programs. Touchscreen PDAs, including Windows Mobile devices, may have a detachable stylus to facilitate making selections. The user interacts with the device by tapping the screen to select buttons or issue commands, or by dragging a finger or the stylus on the screen to make selections or scroll. Typical methods of entering text on touchscreen PDAs include:

- A virtual keyboard, where a keyboard is shown on the touchscreen. Text is entered by tapping the on—screen keyboard with a finger or stylus.
- An external keyboard connected via USB, Infrared port, or Bluetooth. Some users may choose a chorded keyboard for one—handed use.
- Handwriting recognition, where letters or words are written on the touchscreen, and the PDA converts the input to text. Recognition and computation of handwritten horizontal and vertical formulas, such as "1 + 2 =", may also be a feature.
- Stroke recognition allows the user to make a predefined set of strokes on the touchscreen, sometimes in a special input area, representing the various characters to be input. The strokes are often simplified character shapes, making them easier for the device to recognize. One widely—known stroke recognition system is Palm's Graffiti).

Connectivity and Synchronization

Some early PDAs connected to a user's personal computer via serial ports or another proprietary connection, they today connect via a USB cable. Older PDAs from the 90s to 2006 typically had an IrDA (infrared) port allowing short—range, line—of—sight wireless communication. Few current models use this technology, as it has been supplanted by Bluetooth and Wi—Fi.

Most modern PDAs have Bluetooth (a popular wireless protocol for mobile devices) to be used to connect keyboards, headsets, GPS receivers, and other nearby accessories. It's also possible to transfer files between PDAs that have Bluetooth. Many modern PDAs have Wi—Fi wireless network connectivity to connect to Wi—Fi hotspots. All smartphones, and some other modern PDAs like the Apple iPod touch, can connect to Wireless Wide Area Networks which provided by cellular telecommunications companies.

Most PDAs can synchronize their data with applications on a user's personal

computer. This allows the user to update contacts, schedules, or other information on their computer, using software such as Microsoft Outlook, and have that same data transferred to PDA — or transfer updated information from the PDA back to the computer. Synchronization eliminates the need for the user to update their data in two places, it also prevents the loss of information stored on the device if it is lost, stolen, or destroyed. When the PDA is repaired or replaced, it can be "re—synced" with the computer, restoring the user's data. The synchronization software may be part of the computer's operating system, or provided with the PDA, or sold separately by a third party.

Questions:

1) What functions does a PDA perform?
2) Could you describe the main parts of a PDA?
3) Could you say something about PDA's input methods ?
4) What does ***TFT*** stand for in this article?
5) What does the term ***Graffiti*** refer to in this article?

Passage 2 ARM

What is ARM?

ARM is a 32-bit RISC processor architecture developed by ARM Limited, which accounts for over 75% of all 32-bit embedded CPUs, making it one of the most prolific 32-bit architectures in the world. ARM CPUs are found in all corners of consumer electronics, from portable devices (PDAs, mobile phones, media players, handheld gaming units, and calculators) to computer peripherals (hard drives, desktop routers). Important branches in this family include Marvell's XScale and the Texas Instruments OMAP series.

The Evolution of ARM Cores

The ARM design was started in 1983 as a development project at Acorn Computers Ltd. ARM1 was completed by April 1985 and ARM2 come in the following year. The ARM2 featured a 32-bit data bus, a 26-bit address space giving a 64 MB address range and sixteen 32-bit registers. One of these registers served as the (word aligned) program counter with its top 6 bits and lowest 2 bits holding the processor status flags.

The ARM2 was possibly the simplest useful 32-bit microprocessor in the world, with only 30,000 transistors. Much of this simplicity comes from not having microcode, and like most CPUs of the day, not including any cache. This simplicity led to its low power usage, while performing better than the Intel 80286. A successor, ARM3, was produced with a 4KB cache, which further improved performance.

In the late 1980s Apple Computer started working with Acorn on newer versions of the ARM core. The work was so important that Acorn spun off the design team in 1990 into a new company called Advanced RISC Machines Ltd.. For this reason, ARM is sometimes expanded as Advanced RISC Machine instead of Acorn RISC Machine.

Advanced RISC Machines became ARM Ltd when its parent company, ARM Holdings plc, floated on the London Stock Exchange and NASDAQ in 1998.

This work would eventually turn into the ARM6. The first models were released in 1991, and Apple used the ARM6-based ARM 610 as the basis for their Apple Newton PDA. In 1994, Acorn used the ARM 610 as the main CPU in their RISC PC computers.

The core has remained largely the same size throughout these changes. ARM2 had 30,000 transistors, while the ARM6 grew to only 35,000. The idea is that the original design Manufacturer combines the ARM core with a number of optional parts to produce a complete CPU, one that can be built on old semiconductor fabs and still deliver lots of performance at a low cost.

ARM's business has always been to sell IP cores, which licensees use to create microcontrollers and CPUs based on this core. The most successful implementation has been the ARM7TDMI with hundreds of millions sold in almost every kind of microcontroller equipped device.

DEC licensed the architecture and produced the StrongARM. At 233 MHz this CPU drew only 1 watt of power (more recent versions draw far less). This work was later passed to Intel as a part of a lawsuit settlement, and Intel took the opportunity to supplement their aging 1960 line with the StrongARM. Intel later developed its own high performance implementation known as XScale which it has since sold to Marvell.

The common architecture supported on smartphones, PDAs and other handheld devices is ARMv4. XScale and ARM926 processors are ARMv5TE, and are now more numerous in high-end devices than the StrongARM, ARM925T and ARM7TDMI based ARMv4 processors.

Table1 The evolution of ARM cores

Core	Arch	Implemen-tation	Cache I/D/MMU	MIPS@MHz	In Application
ARM1	ARMv1	ARM1	None		ARM Evaluation System second processor for BBC Micro
ARM2	ARMv2	ARM2	None	4 MIPS @ 8MHz	AcornArchimedes,Chessmachine
	ARMv2a	ARM250	None, MEMC1a	7 MIPS @ 12MHz	Acorn Archimedes
ARM3	ARMv2a	ARM2a	4K unified	12 MIPS @ 25MHz	Acorn Archimedes
ARM6	ARMv3	ARM610	4K unified	28 MIPS @ 33MHz	Acorn Risc PC 600,Apple Newton
ARM7	ARMv3	ARM700	8KB unified	40MHz	Acorn Risc PC 700
		ARM710a	8KB unified	40MHz	Apple eMate 300
ARM7TDMI	ARMv4T	ARM7TDMI (-S)	None	15 MIPS @ 16.8 MHz	Game Boy Advance,Nintendo DS,iPod
		ARM710T	8KB unified, MMU	36 MIPS @ 40 MHz	Acorn Risc PC 700, Psion 5 series, Apple eMate 300
		ARM720T	8KB unified, MMU	60 MIPS @ 59.8 MHz	Zipit
		ARM740T	MPU		
	ARMv5TEJ	ARM7EJ-S	None		
StrongARM	ARMv4	SA-110	16KB/16KB	200MHz	
ARM8	ARMv4				
ARM9TDMI	ARMv4T	ARM9TDMI	None		
		ARM920T	16KB/16KB, MMU	200 MIPS @ 180 MHz	Armadillo,GP32,GP2X(first core), Tapwave Zodiac(Motorola i. MX1)
		ARM922T	8KB/8KB, MMU		
		ARM940T	4KB/4KB, MPU		GP2X(second core)
ARM9E	ARMv5TE	ARM946E-S	variable, tightly coupled memories, MPU		Nintendo DS,Nokia N-Gage Conexant 802.11 chips
	ARM966E-S	no cache, TCMs		ST Micro STR91xF	
	ARM968E-S	no cache, TCMs			

续表

ARM9E	ARMv5TEJ	ARM926EJ-S	variable, TCMs, MMU	220 MIPS @ 200 MHz	Mobile phones: Sony Ericsson (K, series), Siemens and Benq (x65series and newer), Texas Instruments OMAP1710
	ARMv5TE	ARM996HS	no cache, TCMs, MPU		
ARM10E	ARMv5TE	ARM1020E	32KB/32KB, MMU		
		ARM1022E	16KB/16KB, MMU		
	ARMv5TEJ	ARM1026EJ-S	variable, MMU or MPU		
XScale	ARMv5TE	80200/IOP310/IOP315			
		80219		400/600MHz	Thecus N2100
		IOP321		600 BogoMips @ 600 MHz	Iyonix
		IOP33x			
		IOP34x	32KB/32KB L1, 512K L2, MMU		
		PXA210/PXA250			Zaurus SL-5600
		PXA255	32KB/32KB, MMU	400 BogoMips @ 400 MHz	Gumstix, Palm Tungsten E2
		PXA26x		up to 400 MHz	Palm Tungsten T3
		PXA27x		800 MIPS @ 624 MHz	HTC Universal, Zaurus SL-C1000, 3000, 3100, 3200, Dell Axim x30, x50 and x51 series
		PXA800(E)F			
		Monahans		1000 MIPS @ 1.25 GHz	Mavell PXA300/PXA310/PXA320, Max frequency : PXA300 @ 624Mhz, PXA310/PXA320@806Mhz
		PXA900			Blackberry 8700, Blackberry Pearl (8100)
		IXC1100			
		IXP2400/IXP2800			
		IXP2850			

续表

XScale	ARMv5TE	IXP2325/ IXP2350			
		IXP42x			NSLU2
		IXP460/ IXP465			
ARM11	ARMv6	ARM1136J (F)-S	variable, MMU	@ 532-665 MHz (i. MX31 SoC)	Nokia N93, Zune, Nokia N800
	ARMv6T2	ARM1156T2 (F)-S	variable, MPU		
	ARMv6KZ	ARM1176JZ (F)-S	variable, MMU+ TrustZone		
	ARMv6K	ARM11 MPCore	variable, MMU		
Cortex	ARMv7-A	Cortex-A8	variable (L1 +L2), MMU +TrustZone	up to 2000 (2.0 DMIPS/ MHz from 600 Mhz to greater than 1 GHz)	Texas Instruments OMAP3
		Cortex-A9			
		Cortex-A9 MPCore			
	ARMv7-R	Cortex-R4 (F)	variable cache, MMU optional	600 DMIPS	Broadcom is a user
	ARMv7-M	Cortex-M3	no cache, (MPU)	120 DMIPS @ 100MHz	Luminary Micro microcontroller family
	ARMv6-M	Cortex-M0			
		Cortex-M1			
	ARMv7-ME	Cortex-M4	Optional 8 region MPU with sub regions and background region	1.25 DMIPS/ MHz	

The ARM Business Model

ARM Ltd. does not manufacture and sell CPU devices based on their own designs,

but rather, licenses the processor architecture to interested parties (see Figure 1). ARM offers a variety of licensing terms, varying in cost and deliverables. To all licensees, ARM provides an integratable hardware description of the ARM core, as well as complete software development toolset (compiler, debugger, SDK), and the right to sell manufactured silicon containing the ARM CPU. Fabless licensees, who wish to integrate an ARM core into their own chip design, are usually only interested in acquiring a ready-to-manufacture verified IP core. For these customers, ARM delivers a gate netlist description of the chosen ARM core, along with an abstracted simulation model and test programs to aid design integration and verification. More ambitious customers, including integrated device manufacturers (IDM) and foundry operators, choose to acquire the processor IP in synthesizable RTL (Verilog) form. With the synthesizable RTL, the customer has the ability to perform architectural level optimizations and extensions. This allows the designer to achieve exotic design goals not otherwise possible with an unmodified netlist (high clock speed, very low power consumption, instruction set extensions, etc.). While ARM does not grant the licensee the right to resell the ARM architecture itself, licensees may freely sell manufactured product (chip devices, evaluation boards, complete systems, etc.). Merchant foundries can be a special case: not only are they allowed to sell finished silicon containing ARM cores, they generally hold the right to remanufacture ARM cores for other customers.

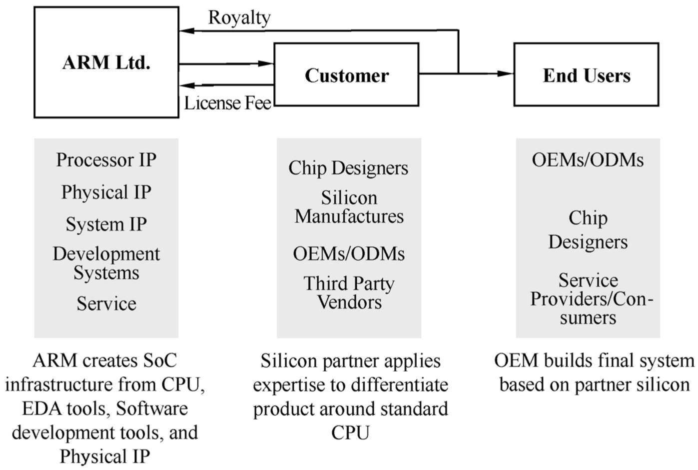


Figure 1 ARM Business Model

Many semiconductor firms hold ARM licenses: Analog Devices, Atmel, Broadcom, Cirrus Logic, Freescale, Fujitsu, Intel, IBM, Infineon Technologies, Nintendo, NXP Semiconductors, OKI, Samsung, Sharp, STMicroelectronics, Texas Instruments and

VLSI are some of the many companies who have licensed the ARM in one form or another. Although ARM's license terms are covered by NDA, within the IP industry, ARM is widely known to be among the most expensive CPU cores. A single customer product containing a basic ARM core can incur a one-time license fee in excess of $200,000. Where significant quantity and architectural modification are involved, the license fee can exceed $10 M.

With the diversity of ARM IP and the broad ecosystem of supporting silicon and software for ARM-based solutions, the world's leading Original Equipment Manufacturers (OEMs) use ARM technology in a wide variety of applications ranging from mobile handsets and digital set top boxes to car braking systems and network routers. Today ARM technology is used in more than 95% of the world's mobile handsets and over one-quarter of all electronic devices.

Questions:

1) What does the term ***ARM*** stand for?
2) For what application are ARM processors mainly used?
3) Could you describe the ARM business model?
4) What do you think is the reason for ARM's popularity?
5) What does the term ***netlist*** mean in this article?

Passage 3 Embedded OS

Embedded operating systems have some features that distinguish them from real-time and general-purpose operating systems. But the definition of an "embedded operating system" is probably even more ambiguous than that of an RTOS, and they come in a zillion different forms. But you'll recognize one when you see one, although the boundary between general-purpose operating systems and embedded operating systems is not sharp, and is even becoming more blurred all the time.

Embedded systems are being installed in tremendous quantities (an order of magnitude more than desktop PCs!): they control lots of functions in modern cars; they show up in household appliances and toys; they control vital medical instrumentation; they make remote controls and GPS (Global Position Systems) work; they make your portable phones work; etc..

The simplest classification between different kinds of embedded operating systems

is as follows:

1. High-end embedded OS. These systems are often downsized derivatives of an existing general purpose OS, but with much of the "ballast" removed. Linux has given rise to a large set of such derivatives, because of its highly modular structure and the availability of source code. Examples are: routers, switches, personal digital assistants, set top boxes.

2. Deeply embedded OS. These OSes must be really very small, and need only a handful of basic functions. Therefore, they are mostly designed from the ground up for a particular application. Two typical functions deeply embedded systems (used to) lack are high-performance graphical user interfacing or network communication. Examples are: automotive controls, digital cameras, portable phones.

The most important features that make an OS into an embedded OS are:

1. Small footprint. Designers are continuously trying to put more computing power in smaller housings, using cheaper CPUs, with on-board digital and/or analog IO; and they want to integrate these CPUs in all kinds of small objects. A small embedded OS also often uses only a couple of kilobytes of RAM and ROM memory.

2. The embedded system should run for years without manual intervention. This means that the hardware and the software should never fail. Hence, the system should preferably have no mechanical parts, such as floppy drives or hard disks. Not only because mechanical parts are more sensitive to failures, but they also take up more space, need more energy, take longer to communicate with, and have more complex drivers (e. g. , due to motion control of the mechanical parts).

3. Many embedded systems have to control devices that can be dangerous if they don't work exactly as designed. Therefore, the status of these devices has to be checked regularly. The embedded computer system itself, however, is one of these critical devices, and has to be checked too! Hence, one often sees hardware watchdogs included in embedded systems. These watchdogs are usually retriggerable monostable timers attached to the processor's reset input. The operating system checks within specified intervals whether everything is working as desired, for example by examining the contents of status registers. It then resets the watchdog. So, if the OS doesn't succeed in resetting the timer, that means that the system is not functioning properly and the timer goes off, forcing the processor to reset.

If something went wrong but the OS is still working (e. g. , a memory protection error in one of the tasks) the OS can activate a software watchdog, which is nothing else but an interrupt that schedules a service routine to handle the error. One important job

of the software watchdog could be to generate a core dump, to be used for analysis of what situations led to the crash.

4. A long autonomy also implies using as little power as possible: embedded systems often have to live a long time on batteries (e.g., mobile phones), or are part of a larger system with very limited power resources (e.g., satellites).

5. If the system does fail despite its designed robustness (e.g., caused by a memory protection fault), there is usually no user around to take the appropriate actions. Hence, the system itself should reboot autonomously, in a "safe" state, and "instantly" if it is supposed to control other critical devices. Compare this to the booting of your desktop computer, which needs a minute or more before it can be used, and always comes up in the same default state。

6. It should be as cheap as possible. Embedded systems are often produced in quantities of several thousands or even millions. Decreasing the unit price even a little bit boils down to enormous savings.

7. Some embedded systems are not physically reachable anymore after they have been started (e.g., launched satellites) in order to add software updates. However, more and more of them can still be accessed remotely. Therefore, they should support dynamic linking: object code that did not exist at the time of start is uploaded to the system, and linked in the running OS without stopping it.

Some applications require all features of embedded and real-time operating systems. The best-known examples are mobile phones and (speech-operated) handheld computers ("PDA"s): they must be small, consume little power, and yet be able to execute advanced signal processing algorithms, while taking up as little space as possible.

The above-mentioned arguments led embedded OS developers to design systems with the absolute minimum of software and hardware. Roughly speaking, developers of general purpose and real-time operating systems approach their clients with a "*Hey, look how much we can do*!" marketing strategy; while EOS developers say "*Hey, look how little we need to do what you want*!" Hence, embedded systems often come without a memory management unit (MMU), multi-tasking, a networking "stack" or file systems. The extreme is one single monolithic program on the bare processor, thus completely eliminating the need for any operating system at all.

Taking out more and more features of a general-purpose operating system makes its footprint smaller and its predictability higher. On the other hand, adding more features to an EOS makes it look like a general purpose OS. Most current RTOS and EOS operating systems are expanding their ranges of application, and cover more of the full

"feature spectrum".

Questions:

1) How are embedded operating systems classified?

2) Can you tell something about Linux?

3) Why hardware watchdogs are often seen included in embedded systems?

4) What applications require all features of embedded and real-time operating systems?

5) Can you give the reasons why general-purpose OSes and embedded OSes take different marketing strategy?

Unit 10

Electronic Instruments & Measurement

Lesson 28　Signal Sources

The Complete Measurement System

An acquisition instrument—usually an oscilloscope or logic analyser—is probably the first thing that comes to mind when you think about making electronic measurements. But these tools can only make a measurement when they are able to acquire a signal of some kind. And there are many instances in which no such signal is available unless it is externally provided.

A strain gauge amplifier, for example, does not produce signals; it merely increases the power of the signals it receives from a sensor. Similarly, a multiplexer on a digital address bus does not originate signals; it directs signal traffic from counters, registers, and other elements. But inevitably it becomes necessary to test the amplifier or multiplexer before it is connected to the circuit that feeds it. In order to use an acquisition instrument to measure the behavior of such devices, you must provide a stimulus signal at the input.

To cite another example, engineers must characterize their emerging designs to ensure that the new hardware meets design specifications across the full range of operation and beyond. This is known as margin or limit testing. It is a measurement task that requires a complete solution; one that can generate signals as well as make measurements.

The toolset for digital design characterization differs from its counterpart in analog/mixed signal design, but both must include stimulus instruments and acquisition instruments. The signal source, or signal generator, is the stimulus source that pairs with an acquisition instrument to create the two elements of a complete measurement solution.

What is a Signal Source?

A signal source is nothing less than the cornerstone of almost any instrumentation setup used in hardware design, debug, or evaluation projects [1]. It is a key engineering tool. It is an essential troubleshooting aid for the technician. It is a surrogate for an automotive ignition pulse, a heart pacemaker, or a guided missile's gyro output. Second

only to the ubiquitous DMM, signal sources are perhaps the most universal class of electronic test instruments [2].

Unless you're working with a purely DC circuit, your circuit is likely to require some kind of AC [3] stimulus signal as you evaluate components, functional blocks, and subsystems. The waveform from the signal source emulates a signal coming in from the outside world, such as a sensor output. Similarly, it can be used as a stand-in for waveforms that will appear in as-yet-unavailable parts of the circuit design.

Interestingly, the signal source's job is not simply to provide an "ideal" waveform. Often the instrument must add known, repeatable amounts and types of distortion (or errors) to the signal it delivers. This characteristic is one of the signal source's strongest virtues, since it is often impossible to create predictable distortion exactly when and where it's needed using only the circuit itself. The response of the unit-under-test (UUT) in the presence of these distorted signals reveals its ability to handle stresses that fall outside the normal performance envelope.

A stimulus signal may take the form of a low-distortion sine wave, a stream of logic pulses, a high-frequency radio carrier wave, a mobile telephone transmission, and many other formats. Traditionally, the task of producing these diverse waveforms has been filled by separate, dedicated signal sources, from ultra-pure audio sine-wave generators to multi-GHz RF signal generators. While there are many commercial solutions, the user must often custom-design or modify a signal source for the project at hand. It can be very difficult to design an instrumentation-quality signal generator, and of course, the time spent designing ancillary test equipment is a costly distraction from the project itself.

Fortunately, digital sampling technology and signal processing techniques have brought us a solution that answers almost any kind of signal generation need with just one type of instrument—the arbitrary generator.

Types of Digital Signal Sources

Broadly divided into arbitrary waveform generators (AWG), arbitrary function generators (AFG), and data or pattern generators (DG), digital signal sources span the whole range of signal-producing needs. Each of these types has its unique strengths:

AWG: Whether you want a data stream shaped by a precise Lorentzian pulse for disk-drive characterization, or a complex modulated RF signal to test a GSM- or CDMA-based telephone handset, the AWG can produce any waveform you can imagine. You can use a variety of methods—from mathematical formulae to "drawing" the waveform-

to create the needed output.

AFG: Typically this instrument offers fewer waveform variations, but with excellent stability and fast response to frequency changes. If the UUT requires the classic "sine and square" waveforms (to name a few) and the ability to switch almost instantly between two frequencies, the AFG is the right tool. An additional virtue is the AFG's low cost, which makes it very attractive for applications that do not require an AWG's versatility.

DG: This third type of signal source meets the special stimulus needs of digital devices that require long, continuous streams of binary data, with specific information content and timing characteristics.

Signal Generation Techniques

There are several ways to create waveforms with a signal source. The choice of methods depends upon the information available about the DUT and its input requirements; whether there is a need to add distortion or error signals, and other variables. Modern high-performance signal sources offer at least three ways to develop waveforms:

- **Simulation**: "Building" an event or sequence of events, based on a specific waveform definition (often from a simulator or a library of waveforms)
- **Replication**: Capturing an existing signal on an oscilloscope and sending it to the signal source for reproduction (Figure 28.1)
- **Substitution**: Creating and/or modifying a defined signal to substitute for a signal from unavailable circuitry (Figure 28.1)

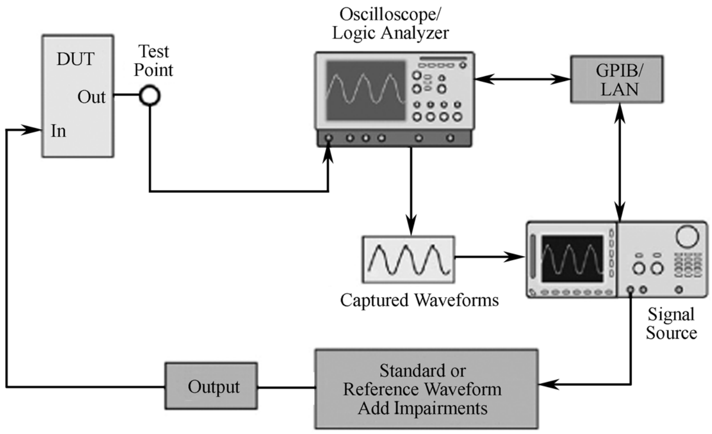


Figure 28.1 Signal sources can use standard, user-created or captured waveforms, adding impairments where necessary for special test applications

Basic Signal Source Applications

Signal sources have hundreds of different applications but in the electronic measurement context they fall into three basic categories: verification, characterization, and stress/margin testing. Some representative applications include:

- **Verification**-Analyzing Digital Modulation: Wireless equipment designers developing new transmitter and receiver hardware must simulate baseband I&Q signals—with and without impairments—to verify conformance with emerging and proprietary wireless standards. Some high-performance arbitrary waveform generators can provide the needed low-distortion, high-resolution signals at rates up to 1 gigabit per second (1 Gbps), with two independent channels, one for the "I" phase and one for the "Q" phase.
- **Characterization**-Testing D/A and A/D Converters: Newly-developed digital-to-analog converters (DAC) and analog-to-digital converters (ADC) must be exhaustively tested to determine their limits of linearity, monotonicity, and distortion. A state-of-the-art [4] AWG can generate simultaneous, in-phase analog and digital signals to drive such devices at speeds up to 1 Gbps.
- **Stress/Margin Testing**-Stressing Communication Receivers: Engineers working with serial data stream architectures (commonly used in digital communications buses and disk drive amplifiers) need to stress their devices with impairments, particularly jitter and timing violations. Advanced signal sources save the engineer untold hours of calculation by providing efficient built-in jitter editing and generation tools. These instruments can shift critical signal edges as little as 0.3 ps.

Signal Source Hardware Architecture: The Arbitrary Waveform Generator

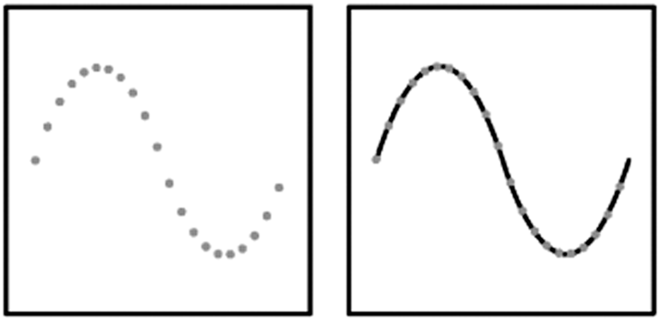
Fundamentally, an Arbitrary Waveform Generator (AWG) is a sophisticated "playback" system that delivers waveforms based on stored digital data that describes the constantly changing voltage levels of an AC signal. It is a tool whose block diagram is deceptively simple.

To understand the AWG, first it's necessary to grasp the broad concepts of digital sampling. Digital sampling is exactly what its name implies: defining a signal using samples, or data points, that represent a series of voltage measurements along the slope of the waveform. These samples may be determined by actually measuring a waveform with an instrument such as an oscilloscope, or by using graphical or mathematical techniques. Figure 28.2a depicts a series of sampled points. All of the points are

sampled at uniform time intervals, even though the curve makes their spacing appear to vary. In an AWG, the sampled values are stored in binary form in a fast RAM.

With the stored information, the signal can be reconstructed at any time by reading back the memory locations and feeding the data points through a DAC. Figure 28. 2b depicts the result. Note that the AWGs output circuitry filters between the points to connect the dots and create a clean, uninterrupted waveform shape. The UUT does not "see"these dots as discrete points, but as a continuous analog waveform. Figure 28. 3 is a simplified block diagram of an AWG that implements these operations.

As always, there are some terms and conditions that affect the fidelity of the signal reproduction. Chief among these is the Nyquist Sampling Theorem, which states that the sampling frequency must be at least twice that of the highest frequency component of the sampled signal. To sample a 1 MHz signal, for instance, it is necessary to acquire points at a frequency of at least 2 MS/s. Although the theorem is usually cited as a guideline for acquisition (as with an oscilloscope), its pertinence to AWGs is clear: stored waveforms must have enough points to faithfully retrace the details of the desired signal.



a b

Figure 28. 2 AWGs use stored digital samples to construct a waveform

An AWG can take these points and read them out of memory at any frequency within its specified limits. If a set of stored points conforms to the Nyquist Theorem and describes a sine wave, then the AWG's output will be a sine wave. However, there is a finite maximum frequency, or sample rate, at which the instrument can operate. This is usually specified in terms of megasamples or gigasamples per second.

Today's fastest AWGs can achieve 2. 6 GS/s. Other AWG hardware characteristics, particularly vertical resolution and memory depth, are just as important as sample rate. Vertical resolution (amplitude) expresses the voltage-measuring precision of the

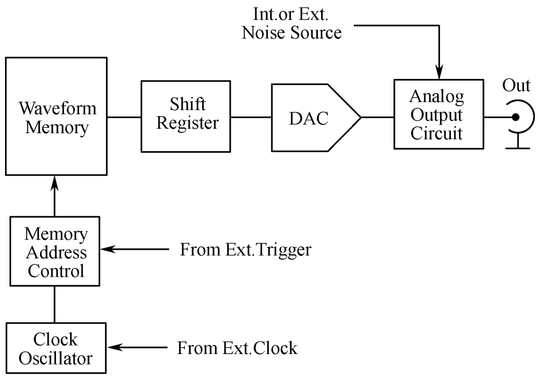


Figure 28.3　AWG block diagram (simplified)

sample points described above. Resolution pertains to the binary word width, in bits, of the instrument's DAC, with more bits equating to higher resolution. While "more is better", higher-frequency AWGs usually have lower resolution — 8 or 10 bits — than general-purpose AWGs offering 12 or 14 bits.

An AWG with 10-bit resolution provides 1024 sample levels spread across the full voltage range of the instrument. If, for example, this 10-bit AWG has a total voltage range of 2 Vp-p, each sample represents a step of approximately 2 mV — the smallest increment the instrument could deliver, assuming it is not constrained by other factors in its architecture.

Memory depth in an AWG plays a key role in the instrument's flexibility. More (deeper) memory provides either of two benefits:

1) More cycles of the desired waveform can be stored. This is useful because it reduces the number of "endpoints". An endpoint is the last memory location occupied by the waveform, after which the AWG must wrap around and return to the beginning in order to continue to produce the output signal. There are unavoidable errors that occur at this transition, so it is desirable to minimize the number of endpoints.

2) More waveform detail can be stored. Complex waveforms have high-frequency information in their pulse edges and transients. It is difficult to interpolate these fast transitions as we did with the simple, predictable sine wave. To faithfully reproduce a complex signal, the available waveform memory capacity must be used to store more transitions and fluctuations rather than more cycles of the signal.

Today's state-of-the-art AWGs offer up to 8-Msample memory depth. When combined with the high sample rate that some top models deliver, these instruments can store and reproduce complex RF waveforms, even including pseudo-random bit streams

for use in physical-layer [5] testing of network equipment. Similarly, these fast AWGs with deep memory can generate very brief digital pulses and transients.

New Words

acquisition [ˌækwiˈziʃn] *n.* 获取,采集
originate [əˈridʒineit] *vt.* 引起 *vi.* 起源
margin [ˈmɑːdʒin] *n.* 边际
cornerstone [ˈkɔːnəstəʊn] *n.* 基础
setup *n.* 设置
evaluation [iˌvæljuˈeiʃn] *n.* 评价,计算
project [ˈprɔdʒekt] *n.* 工程,项目
troubleshooting *n.* 故障排除
surrogate [ˈsʌrəgit] *n.* 代用品,代理人
automotive [ɔːtəˈməutiv] *adj.* 汽车的
ignition [igˈniʃn] *n.* 点火
pacemaker [ˈpeisˌmeikə] *n.* 领跑者,起搏器
gyro [ˈdʒaiərəu] *n.* 陀螺仪
ubiquitous [juːˈbikwitəs] *adj.* 普遍存在的
subsystem [ˈsʌbˌsistim] *n.* 子系统
stand-in *n.* 替身
distortion [disˈtɔːʃn] *n.* 扭曲,失真
ancillary [ænˈsiləri] *adj.* 辅助的,副的
characterization [ˌkæriktəraiˈzeiʃn] *n.* 描述,表征
impairment [imˈpɛəmənt] *n.* 损害
conformance [kənˈfɔːməns] *n.* 顺应,一致
proprietary [prəˈpraiətəri] *adj.* 专有的,所有的
untold [ˌʌnˈtəuld] *adj.* 未说过的,未透露的,数不清的
built-in [ˈbiltˈin] *adj.* 内置的
spacing [ˈspeisiŋ] *n.* 间隔,间距
pertinence [ˈpəːtinəns] *n.* 相关性,针对性
constrained [kənˈstreind] *adj.* 受约束的
endpoint [ˈendpɔint] *n.* 端点,终点
pseudo [ˈsjuːdəu] *adj.* 假的,冒充的
brief [briːf] *adj.* 短暂的

Phrases & Expressions

as yet 到目前为止
to name a few 仅举几例

Technical Terms

oscilloscope [ə'siləskəup] *n*. 示波器
multiplexer ['mʌltiˌpleksə] *n*. 多路复用器
stimulus ['stimjuləs] *n*. 激励信号
instrumentation [ˌinstrumen'teiʃn] *n*. 仪器测量
debug [diː'bʌg] *v*. 调试
gigabit ['dʒigəbit] *n*. 吉比特
linearity [ˌlini'æriti] *n*. 线性
monotonicity [ˌmɔnətə'nisiti] *n*. 单调性
jitter ['dʒitə] *n*. 抖动
slope [sləup] *n*. 斜率
transient ['trænʃənt] *n*. 暂态过程 *adj*. 短暂的
state-of-the-art *n*. 技术水平，工艺水平
in-phase *n*. 同相
logic analyzer 逻辑分析仪
stimulus signal 激励信号
guided missile 导弹
design specification 设计规格
full range 满量程
desired signal 期望信号
vertical resolution 垂直分辨率
DMM *abbr*. Digital Multimeter 数字多用表
UUT *abbr*. Unit Under Test 被测单元
AWG *abbr*. Arbitrary Waveform Generator 任意波形发生器
AFG *abbr*. Arbitrary Function Generator 任意函数发生器
DG *abbr*. Data Generator 数据发生器

Notes

1. 此句可译为:信号源是用于硬件设计、调试和评价工程中的几乎任何一种测试配置的基础。nothing less than 是“不亚于，完全”的意思。
2. 此句可译为:信号源也许是在电子测试仪器中仅次于无处不在的数字多用表的最为通用的一类仪器。second only to 是“仅次于”的意思。
3. 通常情况下,AC 是指在 0 V 上下交替改变方向的信号(即交流信号)。在有些情况下,AC 可以指任何变化的信号(如幅度在+1～ +3 V 之间变化的信号)。
4. state of the art 是指一种设备、一门技术或一个科学领域在某个特定时代取得的最高发展水平。
5. 物理层(physical layer)是数据通信中的开放系统互连(OSI，Open System Interconnection)参考模型的最底层。OSI 模型由 7 层组成,其余 6 层依次为数据链路层(data link layer)、网络层(network layer)、传输层(transport layer)、会话层(session layer)、表示层(presentation layer)和应用层 (application layer)。OSI 模型中的每一层执行一项特定的任务。其中,物理层负责位流的传输。

Lesson 29 Oscilloscopes

Nature moves in the form of a sine wave, be it an ocean wave, earthquake, sonic boom [1], explosion, sound through air, or the natural frequency of a body in motion [2]. Energy, vibrating particles and other invisible forces pervade our physical universe. Even light—part particle, part wave—has a fundamental frequency, which can be observed as color. Sensors can convert these forces into electrical signals that you can observe and study with an oscilloscope. Oscilloscopes enable scientists, engineers, technicians, educators and others to “see” events that change over time.

Oscilloscopes are indispensable tools for anyone designing, manufacturing or repairing electronic equipment. In today's fast-paced world, engineers need the best tools available to solve their measurement challenges quickly and accurately. As the eyes of the engineer, oscilloscopes are the key to meeting today's demanding measurement challenges.

The usefulness of an oscilloscope is not limited to the world of electronics. With the proper transducer, an oscilloscope can measure all kinds of phenomena. A transducer is

a device that creates an electrical signal in response to physical stimuli, such as sound, mechanical stress, pressure, light, or heat. A microphone is a transducer that converts sound into an electrical signal.

Oscilloscopes are used by everyone from physicists to television repair technicians. An automotive engineer uses an oscilloscope to measure engine vibrations. A medical researcher uses an oscilloscope to measure brain waves. The possibilities are endless.

The Types of Oscilloscopes

Electronic equipment can be classified into two categories: analog and digital. Analog equipment works with continuously variable voltages, while digital equipment works with discrete binary numbers that represent voltage samples.

Oscilloscopes can be classified similarly-as analog and digital types. For many applications, either an analog or digital oscilloscope will do. However, each type has unique characteristics that may make it more or less suitable for specific applications. Digital oscilloscopes can be further classified into digital storage oscilloscopes (DSOs), digital phosphor oscilloscopes (DPOs) and sampling oscilloscopes.

Analog Oscilloscopes

Fundamentally, an analog oscilloscope works by applying the measured signal voltage directly to the vertical axis of an electron beam that moves from left to right across the oscilloscope screen-usually a cathode-ray tube (CRT). The backside of the screen is treated with luminous phosphor that glows wherever the electron beam hits it. The signal voltage deflects the beam up and down proportionally as it moves horizontally across the display, tracing the waveform on the screen. The more frequently the beam hits a particular screen location, the more brightly it glows.

The CRT limits the range of frequencies that can be displayed by an analog oscilloscope. At very low frequencies, the signal appears as a bright, slow-moving dot that is difficult to distinguish as a waveform. At high frequencies, the CRT's writing speed defines the limit. When the signal frequency exceeds the CRT's writing speed, the display becomes too dim to see. The fastest analog oscilloscopes can display frequencies up to about 1 GHz.

When you connect an oscilloscope probe to a circuit, the voltage signal travels through the probe to the vertical system of the oscilloscope. Figure 29.1 illustrates how an analog oscilloscope displays a measured signal. Depending on how you set the vertical scale (volts/div control), an attenuator reduces the signal voltage and an amplifier

increases the signal voltage.

Next, the signal travels directly to the vertical deflection plates of the CRT. Voltage applied to these deflection plates causes a glowing dot to move across the screen. The glowing dot is created by an electron beam that hits the luminous phosphor inside the CRT. A positive voltage causes the dot to move up while a negative voltage causes the dot to move down. The signal also travels to the trigger system to start, or trigger, a horizontal sweep. Horizontal sweep refers to the action of the horizontal system that causes the glowing dot to move across the screen(Figure 29.1).

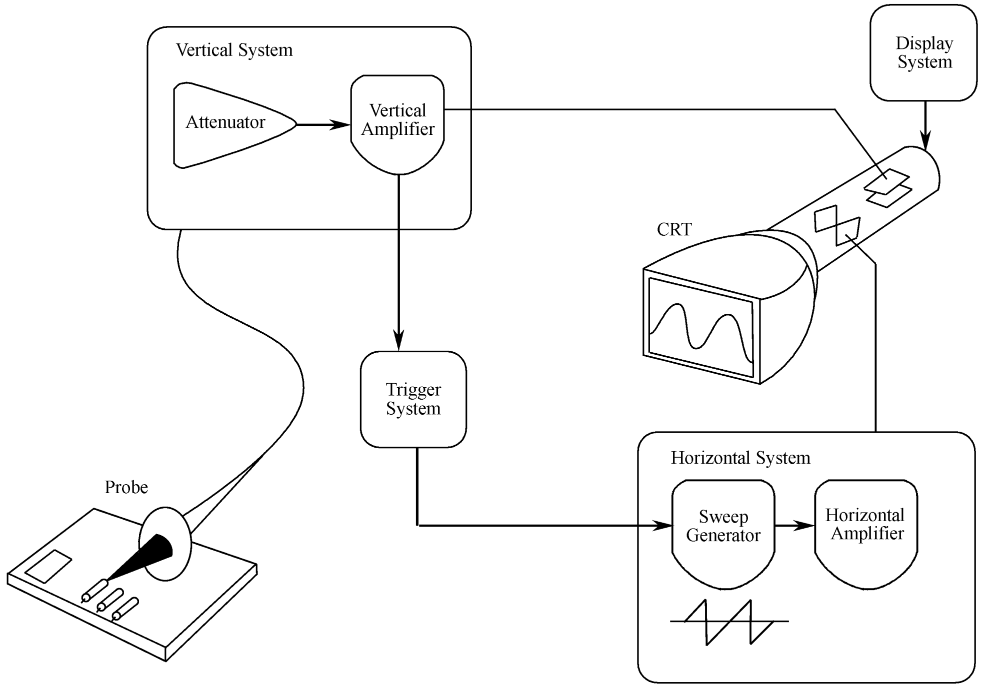


Figure 29.1　An analog oscilloscope

Triggering the horizontal system causes the horizontal time base to move the glowing dot across the screen from left to right within a specific time interval. Many sweeps in rapid sequence cause the movement of the glowing dot to blend into a solid line. At higher speeds, the dot may sweep across the screen up to 500,000 times per second.

Together, the horizontal sweeping action and the vertical deflection action trace a graph of the signal on the screen. The trigger is necessary to stabilize a repeating signal it ensures that the sweep begins at the same point of a repeating signal.

In addition, analog oscilloscopes have focus and intensity controls that can be adjusted to create a sharp, legible display. People often prefer analog oscilloscopes when it is important to display rapidly varying signals in "real time"-or, as they occur. The analog oscilloscope's chemical phosphor-based display has a characteristic known as intensity grading that makes the trace brighter wherever the signal features occur most often. This intensity grading makes it easy to distinguish signal details just by looking at the trace's intensity levels.

Digital Oscilloscopes

In contrast to an analog oscilloscope, a digital oscilloscope uses an ADC to convert the measured voltage into digital information. It acquires the waveform as a series of samples, and stores these samples until it accumulates enough samples to describe a waveform. The digital oscilloscope then re-assembles the waveform for display on the screen. Digital oscilloscopes can be classified into digital storage oscilloscopes (DSOs), digital phosphor oscilloscopes (DPOs), and sampling oscilloscopes.

The digital approach means that the oscilloscope can display any frequency within its range with stability, brightness, and clarity. For repetitive signals, the bandwidth of the digital oscilloscope is a function of the analog bandwidth of the front-end components of the oscilloscope, commonly referred to as the 3 dB point. For single-shot and transient events, such as pulses and steps, the bandwidth can be limited by the oscilloscope's sample rate.

Digital Storage Oscilloscopes

A conventional digital oscilloscope is known as a digital storage oscilloscope (DSO). Its display typically relies on a raster-type screen rather than luminous phosphor. DSOs allow you to capture and view events that may happen only once-known as transients. Because the waveform information exists in digital form as a series of stored binary values, it can be analyzed, archived, printed, and otherwise processed, within the oscilloscope itself or by an external computer. The waveform need not be continuous; it can be displayed even when the signal disappears. Unlike analog oscilloscopes, digital storage oscilloscopes provide permanent signal storage and extensive waveform processing. However, DSOs typically have no real-time intensity grading; therefore, they cannot express varying levels of intensity in the live signal.

Digital Phosphor Oscilloscopes

The digital phosphor oscilloscope (DPO) offers a new approach to oscilloscope architecture. This architecture enables a DPO to deliver unique acquisition and display

capabilities to accurately reconstruct a signal. While a DSO uses a serial-processing architecture (see Figure 29.2) to capture, display and analyze signals, a DPO employs a parallel-processing architecture to perform these functions, as shown in Figure 29.3. The DPO architecture dedicates unique ASIC hardware to acquire waveform images, delivering high waveform capture rates that result in a higher level of signal visualization. This performance increases the probability of witnessing transient events that occur in digital systems, such as runt pulses, glitches and transition errors.

Digital Sampling Oscilloscopes

When measuring high-frequency signals, the oscilloscope may not be able to collect enough samples in one sweep. A digital sampling oscilloscope is an ideal tool for accurately capturing signals whose frequency components are much higher than the oscilloscope's sample rate. This oscilloscope is capable of measuring signals of up to an order of magnitude faster than any other oscilloscope. It can achieve bandwidth and high-speed timing ten times higher than other oscilloscopes for repetitive signals. Sequential equivalent-time sampling oscilloscopes are available with bandwidths to 50 GHz.

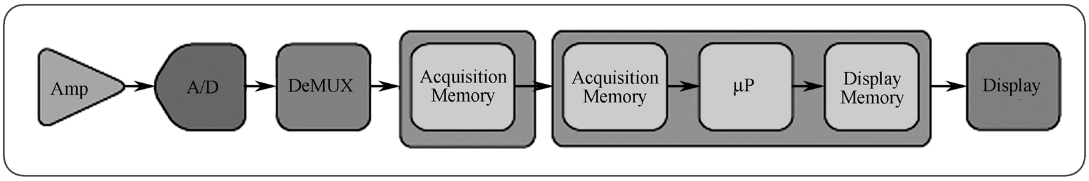


Figure 29.2 The serial-processing architecture of a digital storage oscilloscope (DSO)

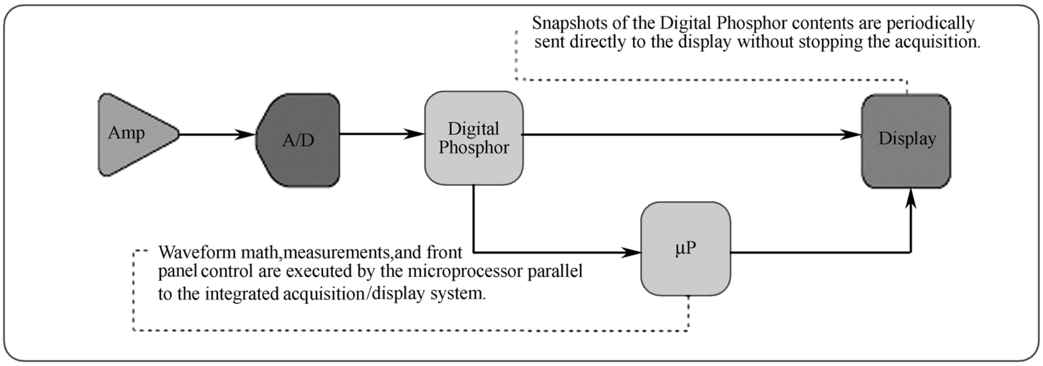


Figure 29.3 The parallel-processing architecture of a digital phospor oscilloscope (DPO)

New Words

vibrating [vai'breitiŋ] *n*. 振动
particle ['pɑːtikl] *n*. 粒子
pervade [pə'veid] *v*. 遍及
indispensable [ˌindi'spensəbl] *adj*. 不可缺少的，绝对必要的
demanding [di'mɑːndiŋ] *adj*. 过分的，苛求的
phosphor ['fɔsfə] *n*. 磷，启明星
backside ['bæksaid] *n*. 背部，后方
luminous ['luːminəs] *adj*. 发光的，明亮的
legible ['ledʒəbl] *adj*. 清晰的，易读的
grading ['ɡreidiŋ] *n*. 分级，归类
front-end *adj*. 前端的，前期的
archive ['ɑːkaiv] *vt*. 存档 *n*. 档案文件

Phrases & Expressions

… refer to… ……指的是……
in contrast to … 与……不同
(be) known as … 称作……

Technical Terms

transducer [trænz'djuːsə] *n*. 传感器，换能器
cathode ['kæθəud] *n*. 阴极
attenuator [ə'tenjueitə] *n*. 衰减器
raster ['ræstə] *n*. 光栅，屏面
single-shot *n*. 单脉冲
sonic boom 声爆
electron beam 电子束
fundamental frequency 基频
CRT *abbr*. Cathode Ray Tube 阴极射线管

Notes

1. 声爆(sonic boom)是一种在以音速或超音速飞行的飞行器前方的冲击波引起的爆炸声。
2. 此句可译为:自然界是以正弦波的形式运动着。不论是海浪、地震、声爆和爆炸,还是空气中的声音,或者是运动物体的固有频率都是如此。

Lesson 30 Logic Analyzers

The logic analyzer has different capabilities than the oscilloscope. The most obvious difference between the two instruments is the number of channels (inputs). Typical digital oscilloscopes have up to four signal inputs. Each channel inputs one digital signal. Some complex system designs require thousands of input channels. Appropriately-scaled logic analyzers are available for those tasks as well.

A logic analyzer measures and analyzes signals differently than an oscilloscope. The logic analyzer doesn't measure analog details. Instead, it detects logic threshold levels. When you connect a logic analyzer to a digital circuit, you're only concerned with the logic state of the signal. A logic analyzer looks for just two logic levels. When the input is above the threshold voltage is said to be"high"or"1;"conversely, the level below the threshold voltage is a"low"or"0."When a logic analyzer samples input, it stores a"1"or a"0"depending on the level of the signal relative to the voltage threshold.

A logic analyzer's waveform timing display is similar to that of a timing diagram found in a data sheet or produced by a simulator. All of the signals are time-correlated, so that setup-and-hold time, pulse width, extraneous or missing data can be viewed. In addition to their high channel count, logic analyzers offer important features that support digital design verification and debugging. Among these are:

• Sophisticated triggering that lets you specify the conditions under which the logic analyzer acquires data.

• High-density probes and adapters that simplify connection to the system under test (SUT).

• Analysis capabilities that translate captured data into processor instructions and correlate it to source code.

Architecture and Operation

The logic analyzer connects to, acquires, and analyzes digital signals. These are the four steps to using a logic analyzer:

1. Probe (connect to the System Under Test-SUT)
2. Setup (clock mode and triggering)
3. Acquire
4. Analyze and display

Figure 30.1 is a simple logic analyzer block diagram. Each block symbolizes several hardware and/or software elements. The block numbers correspond to the four steps listed above.

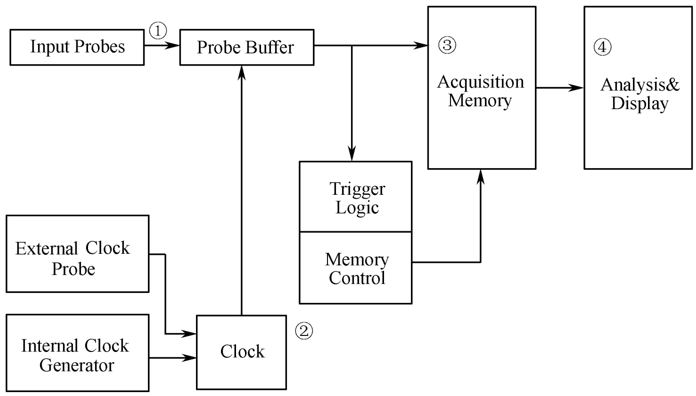


Figure 30.1　Simplified logic analyzer block diagram

Probe

The large number of signals that can be captured at one time by the logic analyzer is what sets it apart from the oscilloscope. The acquisition probes connect to the System Under Test (SUT). The probe's internal comparator is where the input voltage is compared against the threshold voltage (V_{th}), and where the decision about the signal's logic state (1 or 0) is made. The threshold value is set by the user, ranging from TTL levels to, CMOS, ECL, and user-definable.

The probes are capable of acquiring high-quality signals, and have a minimal impact on the SUT. The logic analyzer probe provides a high-quality signal path to the logic analyzer, minimizes electrical loading on the SUT and adapts to the various types of connections on circuit boards and devices [1].

The impedance of the logic analyzer's probes (capacitance, resistance, and inductance) becomes part of the overall load on the circuit being tested. All probes exhibit loading characteristics. The logic analyzer probe should introduce minimal loading on the SUT, and provide an accurate signal to the logic analyzer.

Probe capacitance tends to "roll off" the edges of signal transitions. Why is this important? Because a slower edge crosses the logic threshold of the circuit later, introducing timing errors in the SUT. This is a problem that becomes more severe as clock rates increase. In high-speed systems, excessive probe capacitance can potentially prevent the SUT from working! It is always critical to choose a probe with the lowest possible total capacitance.

Setup (Clock Mode and Triggering)

Logic analyzers are designed to capture data from multi-pin devices and buses. The term "capture rate" refers to how often the inputs are sampled [2]. It is the same function as the time base in an oscilloscope. There are two types of data acquisition:

1) Asynchronous acquisition captures signal timing information. In this mode, a clock internal to the logic analyzer is used to sample data. The faster that data is sampled, the higher will be the resolution of the measurement. There is no fixed timing relationship between the target device and the data acquired by the logic analyzer. This acquisition mode is primarily used when the timing relationship between SUT signals is of primary importance.

2) Synchronous acquisition is used to acquire the "state" of the SUT. A signal from the SUT defines the sample point (when and how often data will be acquired). The signal used to clock the acquisition may be the system clock, a control signal on the bus, or a signal that causes the SUT to change states. Data is sampled on the active edge and it represents thecondition of the SUT when the logic signals are stable. The logic analyzer samples when, and only when, the chosen signals are valid.

Acquisition: Real-time Acquisition Memory

The logic analyzer's probing, triggering, and clocking systems exist to deliver data to the real-time acquisition memory. This memory is the heart of the instrument—the destination for all of the sampled data from the SUT, and the source for all of the instrument's analysis and display.

Logic analyzers have memory capable of storing data at the instrument's sample rate. This memory can be envisioned as a matrix having width and depth.

The instrument accumulates a record of all signal activity until a trigger event or the user tells it to stop. The result is an acquisition—essentially a multi-channel waveform display that lets you view the interaction of all the signals you've acquired, with a very high degree of timing precision.

Both width and depth are key factors in choosing a logic analyzer. Following [3] are some tips to help you determine your channel count and memory depth:

How many signals do you need to capture and analyze? Your logic analyzer's channel count maps directly to the number of signals you want to capture. Digital system buses come in various widths, and there is often a need to probe other signals (clocks, enables, etc.) at the same time the full bus is being monitored. Be sure to consider all the buses and signals you will need to acquire simultaneously.

How much "time" do you need to acquire? This determines the logic analyzer's memory depth requirement, and is especially important for asynchronous acquisition. For a given memory capacity, the total acquisition time decreases as the sample rate increases. For example, the data stored in a 1M memory spans 1 second of time when the sample rate is 1 ms. The same 1M memory spans only 10 ms of time for an acquisition clock period of 10 ns. Acquiring more samples (time) increases your chance of capturing both an error, and the fault that caused the error. When it comes to memory capacity, you can't have too much!

Analysis and Display

The data stored in the real-time acquisition memory can be used in a variety of display and analysis modes. Once the information is stored within the system, it can be viewed in formats ranging from timing waveforms to instruction mnemonics correlated to source code.

The waveform display is a multi-channel detailed view that lets you see the time relationship of all the captured signals, much like the display of an oscilloscope. The waveform display is commonly used in timing analysis, and it is ideal for:

- Diagnosing timing problems in SUT hardware.
- Verifying correct hardware operation by comparing the recorded results with simulator output or data sheet timing diagrams.
- Measuring hardware timing-related characteristics:
  —Race conditions
  —Propagation delays
  —Absence or presence of pulses

• Analyzing glitches

The listing display provides state information in user-selectable alphanumeric form. The data values in the listing are developed from samples captured from an entire bus and can be represented in hexadecimal or other formats. The intent of the listing display is to show the state of the SUT. The listing display lets you see the information flow exactly as the SUT sees it—a stream of data words.

State data is displayed in several formats. The-real-time instruction trace disassembles every bus transaction and determines exactly which instructions were read across the bus. It places the appropriate instruction mnemonic, along with its associated address, on the logic analyzer display.

An additional display, the source code debug display, makes your debug work more efficient by correlating the source code to the instruction trace history. It provides instant visibility of what's actually going on when an instruction executes.

New Words

concerned [kən'sə:nd] *adj*. 有关的

diagram ['daiəgræm] *n*. 图表

mnemonics [ni:'mɔniks] *n*. 助记符

setup ['setəp] *n*. 设置

extraneous [ik'streiniəs] *adj*. 无关的,外来的

envision [in'viʒən] *vt*. 想象,预想

glitch [glitʃ] *n*. 毛刺

alphanumeric [ˌælfənju:'merik] *adj*. 包括文字与数字的

Phrases & Expressions

be concerned with … 对……关心

correspond to … 与……相对应

when it comes to … 谈到……

Technical Terms

threshold ['θreʃhəuld] *n*. 门限

probe [prəub] *n*. 探针,探头

loading ['ləudiŋ] *n*. 载荷，负载
impedance [im'pi:dəns] *n*. 阻抗
asynchronous [ei'siŋkrənəs] *adj*. 异步的
synchronous ['siŋkrənəs] *adj*. 同步的
timing diagram 时序图
pulse width 脉冲宽度
time base 时基
propagation delay 传播延迟
data acquisition 数据采集
SUT *abbr*. System Under Test 被测系统

Notes

1. 此句可译为：逻辑分析仪的探头为逻辑分析仪提供高质量信号通道、尽量减少(由测试引起的)对被测系统的电气载荷、对电路板和器件上各种类型的连接器进行适配。
2. 此句可译为：捕获速率是指对输入信号进行采样的频率。
3. Following 可用作名词、形容词或介词。Following 在这里用作名词，意思为“下列各项”。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) Proper grounding is an important step ______ setting up to take measurements or work ______ a circuit. Proper grounding of the oscilloscope protects you from a hazardous shock and grounding yourself protects your circuits ______ damage.

To ground the oscilloscope means to connect it to an electrically neutral ______(参考点), such as earth ground. Ground your oscilloscope by plugging its three-pronged power cord into an outlet grounded to earth ground.

Grounding the oscilloscope is necessary for safety. If ______(一个高电压)contacts the case of an ungrounded oscilloscope—any part of the case, ______ knobs that appear insulated—it can give you a shock. However, with a properly grounded oscilloscope, the current travels through the grounding path to earth ground ______ through you to earth ground.

Grounding is also necessary for ______(进行精确测量)with your oscilloscope. The

oscilloscope needs to ______(共用)the same ground as any circuits you are testing.

Some oscilloscopes do not require separate connection to earth ground. These oscilloscopes have insulated cases and controls, which keeps any possible shock hazard away from the user.

If you are working with ______(集成电路), you also need to ground yourself. Integrated circuits have tiny ______(导电通路)that can be damaged by ______(静电) that builds up on your body. You can ruin an expensive IC simply by walking across a carpet or taking off a sweater and then touching the ______(引线)of the IC. To solve this problem, wear a grounding strap. This strap safely sends static ______(电荷)on your body to earth ground.

(2) Many engineers look at tasks like troubleshooting and ______(设计验证) as purely "measurement" challenges, and tend to think reflexively of their ______(示波器或者逻辑分析仪) as the whole solution ______ the problem. But these acquisition instruments have an important ______(伙伴) in their work: the ______ (激励) instrument—the signal source.

Stimulus and acquisition instruments together ______ (构成) a complete solution that can drive a ______(被测设备) with complex ______(现实世界信号) and acquire the resulting outputs. The oscilloscope is the ______(工业标准工具) for acquisition. But only with a signal source ______ engineers really control what goes into the device. And that is often a necessity ______ making sense of what comes out of the device.

______(类似地), the signal source makes margin testing and characterization possible. Working with a signal source and an oscilloscope or logic analyzer, engineers can explore the limits of ______(其设计的性能)— introducing deliberate stresses with the source and measuring the results ______ the oscilloscope, or capturing data with the logic analyzer when digital errors occur.

In applications ranging ______ disk drive design to telecommunications conformance testing, the signal source and the acquisition instruments work together as ______(一个完整的测量方案).

2. Translate the following passages into Chinese or English.

1) The term "wave" can be defined as a pattern of varying quantitative values that repeats over some interval of time. Waves are common in nature: sound waves, brain waves, ocean waves, light waves, voltage waves, and many more. All are periodically repeating phenomena. Signal sources are usually concerned with producing electrical (typically voltage) waves that repeat in a controllable manner.

2) Waveforms have many characteristics but their key properties pertain to amplitude, frequency, and phase. The amplitude, frequency, and phase characteristics of a waveform are the building blocks a signal source uses to optimize waveforms for almost any application.

3) A multimeter is an instrument for measuring electricity (volts, amps, ohms) that is widely used and available in numerous shapes and sizes. An analog multimeter displays results by moving a pointer across a printed scale. The movement of the pointer helps to show gradual changes, but is less precise than its digital counterpart. A digital multimeter (DMM) displays the exact measurement on a numeric LCD or LED readout.

4) An oscilloscope is a test instrument that displays electronic signals (waves and pulses) on a screen. It creates its own time base against which signals can be measured.

5) Selecting the right probe for your application is key to attaining the best signal fidelity in your measurements. Active probes provide truer signal reproduction and fidelity for high-frequency measurements.

6)信号幅度的变化不一定是以 0 V 参考地电位为中心的。电路地电位和信号中心电压之差称作“偏置电压”。

7)“差分信号”是指使用两路互补通道的信号。两路互补通道分别以正、反极性传输同一个信号。

8)“上升时间”和“下降时间”是指信号边沿进行状态转换所用的时间。在现代数字电路中,其数值通常是几个纳秒。

9)“带宽”是指通信线路、计算机总线等电气信道的传输能力。在数字线路中,“带宽”是以“位速率”或“字节速率”为单位。在模拟线路中,“带宽”是指最高频率和最低频率之差,并以“赫兹”为单位。

10)“噪声”是一种侵入到电气传输中的无关信号。“噪声”来源可能是附近电线的强电磁信号、匹配欠佳的电气连接和电力线尖峰信号。

Reading Materials

Passage 1 Understanding Waveforms

Basic Waves

Waveforms come in many shapes and forms. Most electronic measurements use one or more of the following wave shapes, often with noise or distortion added:

- Sine waves
- Square and rectangular waves
- Sawtooth and triangle waves
- Step and pulse shapes
- Complex waves

Sine Waves

Sine waves are perhaps the most recognizable wave shape. Most AC power sources produce sine waves. Household wall outlets deliver power in the form of sine waves. And the sine wave is almost always used in elementary classroom demonstrations of electrical and electronic principles. The sine wave is the result of a basic mathematical function—graphing a sine curve through 360 degrees will produce a definitive sine wave image. The damped sine wave is a special case in which a circuit oscillates from an impulse, and then winds down over time. Many applications such as frequency response tests call for a "swept" sine wave—one that increases in frequency over some period of time. In effect, this is a form of frequency modulation.

Square And Rectangular Waves

Square and rectangular waves are basic forms that are at the heart of all digital electronics, and they have other uses as well. A square wave is a voltage that switches between two fixed voltage levels at equal intervals. It is routinely used to test amplifiers, which should be able to reproduce the fast transitions between the two voltage levels (these are the rise and fall times explained earlier). The square wave makes an ideal timekeeping clock for digital systems-computers, wireless telecom equipment, HDTV systems, and more. A rectangular wave has switching characteristics similar to those of a square wave, except that its high and low time intervals are not of equal length.

Sawtooth and Triangle Waves

Sawtooth and triangle waves look very much like the geometric shapes they are

named for. The sawtooth ramps up slowly and evenly to a peak in each cycle, then falls off quickly. The triangle has more symmetrical rise and fall times. These waveforms are often used to control other voltages in systems such as analog oscilloscopes and televisions.

Step and Pulse Shapes

A"step"is simply a waveform that shows a sudden change in voltage, as though a power switch had been turned on. The"pulse" is related to the rectangular wave. Like the rectangle, it is produced by switching up and then down, or down and then up, between two fixed voltage levels. Pulses are inherently binary and therefore are the basic tool for carrying information (data) in digital systems. A pulse might represent one bit of information traveling through a computer. A collection of pulses traveling together creates a pulse train. A synchronized group of pulse trains (which may be transmitted in parallel or serial fashion) makes up a digital pattern.

Note that, while digital data is nominally made up of pulses, rectangles, and square waves, real-world digital waveforms exhibit more rounded corners and slanted edges.

Sometimes, circuit anomalies produce pulses spontaneously. Usually these transient signals occur non-periodically, and have come to be described with the term "glitch". One of the challenges of digital troubleshooting is distinguishing glitch pulses from valid but narrow data pulses. And one of the strengths of certain types of signal sources is their ability to add glitches anywhere in a pulse train.

Complex Waves

In operational electronic systems, waveforms rarely look like the textbook examples explained above. Certain clock and carrier signals are pure, but most other waveforms will exhibit some unintended distortion (a by-product of circuit realities like distributed capacitance, crosstalk, and more) or deliberate modulation. Some waveforms may even include elements of sines, squares, steps, and pulses. Complex waves include:

- Analog modulated, digitally modulated, pulse-width modulated, and quadrature modulated signals
- Digital patterns and formats
- Pseudo-random bit and word streams

In modulated signals, amplitude, phase and/or frequency variations embed lower-frequency information into a carrier signal of higher frequency. The resulting signals may convey anything from speech to data to video. The waveforms can be a challenge to reproduce unless the signal source is specifically equipped to do so.

Analog Modulation

Amplitude modulation (AM) and frequency modulation (FM) are most commonly used in broadcast communications. The modulating signal varies the carrier's amplitude and/or frequency. At the receiving end, demodulating circuits interpret the amplitude and/or frequency variations, and extract the content from the carrier. Phase modulation (PM) modulates the phase rather than the frequency of the carrier waveform to embed the content.

Digital Modulation

Digital modulation, like other digital technologies, is based on two states which allow the signal to express binary data. In amplitude-shift keying (ASK), the digital modulating signal causes the output frequency to switch between two amplitudes; in frequency-shift keying (FSK), the carrier switches between two frequencies (its center frequency and an offset frequency); and in phase-shift keying (PSK), the carrier switches between two-phase settings. In PSK, a"0"is imparted by sending a signal of the same phase as the previous signal, while a"1"bit is represented by sending a signal of the opposite phase. Pulse-width modulation (PWM) is yet another common digital format; it is often used in digital audio systems. As its name implies, it is applicable to pulse waveforms only. With PWM, the modulating signal causes the active pulse width (duty cycle, explained earlier) of the pulse to vary.

Quadrature Modulation

Today's digital wireless communications networks are built on a foundation of quadrature (IQ) modulation technology. Two carriers—an in-phase (I) waveform and a quadrature-phase (Q) waveform that is delayed by exactly 90 degrees relative to the"I" waveform—are modulated to produce four states of information. The two carriers are combined and transmitted over one channel, then separated and demodulated at the receiving end. The IQ format delivers far more information than other forms of analog and digital modulation: it increases the effective bandwidth of the system.

Digital Patterns and Formats

A digital pattern consists of multiple synchronized pulse streams that make up "words"of data that may be 8, 12, 16, or more bits wide. One class of signal source, the digital pattern generator, specializes in delivering words of data to digital buses and processors via parallel outputs. The words in these patterns are transmitted in a steady march of cycles, with the activity for each bit in each cycle determined by the chosen signal format. The formats affect the width of the pulses within the cycles that compose the data streams.

The following list summarizes the most common formats. The first three format explanations assume that the cycle begins with a binary"0"value—that is, a low logic voltage level.

• Non-Return-to-Zero (NRZ): When a valid bit occurs in the cycle, the waveform switches to a"1"and stays at that value until the next cycle boundary.

• Delayed Non-Return-to-Zero (DNRZ): Similar to NRZ, except that the waveform switches to a "1" after a specified delay time.

• Return-to-Zero (RZ): When a valid bit is present, the waveform switches to a "1", then back to a "0"within the same cycle.

• Return-to-One (R1): In effect, the inverse of RZ. Unlike the other formats in this list, R1 assumes the cycle begins with a"1", then switches to a"0"when the bit is valid, then switches back to a"1"before the cycle ends.

Bit Streams

Pseudo-random bit streams (PRBS) and pseudo-random word streams (PRWS) exist to make up for an innate limitation in digital computers: their inability to produce truly random numbers. Yet random events can have beneficial uses in digital systems. For example, perfectly"clean"digital video signals may have objectionable jagged lines and noticeable contours on surfaces that should be smooth. Adding a controlled amount of noise can hide these artefacts from the eye without compromising the underlying information.

To create random noise, digital systems produce a stream of numbers that has the effect of randomness even though the numbers follow a predictable mathematical pattern. These"pseudo-random"numbers are actually a set of sequences repeated at a random rate. The result is a PRBS. A pseudo-random word stream defines how multiple PRBS streams are presented across the signal source's parallel outputs.

PRWS is often used when testing serializers or multiplexers. These elements reassemble the PRWS signal into a serial stream of pseudo-random bits.

Questions:

1) What wave shape do you think is the most recognizable one? Why?

2) What waveform is widely used in digital systems?

3) How many types of complex wave are given in this article? Can you describe some of them?

4) What is a digital pattern?

5) What does ***PRBS*** stand for? Is it identical to PRWS?

Passage 2 Signal Integrity

The Significance of Signal Integrity

The key to any good oscilloscope system is its ability to accurately reconstruct a waveform—referred to as signal integrity. An oscilloscope is analogous to a camera that captures signal images that we can then observe and interpret. Two key issues lie at the heart of signal integrity.

- When you take a picture, is it an accurate picture of what actually happened? Is the picture clear or fuzzy?
- How many of those accurate pictures can you take per second?

Taken together, the different systems and performance capabilities of an oscilloscope contribute to its ability to deliver the highest signal integrity possible. Probes also affect the signal integrity of a measurement system.

Signal integrity impacts many electronic design disciplines. But until a few years ago, it wasn't much of a problem for digital designers. They could rely on their logic designs to act like the Boolean circuits they were. Noisy, indeterminate signals were something that occurred in high-speed designs—something for RF designers to worry about. Digital systems switched slowly and signals stabilized predictably.

Processor clock rates have since multiplied by orders of magnitude. Computer applications such as 3D graphics, video and server I/O demand vast bandwidth. Much of today's telecommunications equipment is digitally based, and similarly requires massive bandwidth. So too does digital high-definition TV. The current crop of microprocessor devices handles data at rates up to 2, 3 and even 5 GS/s (gigasamples per second), while some memory devices use 400-MHz clocks as well as data signals with 200-ps rise times.

Importantly, speed increases have trickled down to the common IC devices used in automobiles, VCRs, and machine controllers, to name just a few applications. A processor running at a 20-MHz clock rate may well have signals with rise times similar to those of an 800-MHz processor. Designers have crossed a performance threshold that means, in effect, almost every design is a high-speed design.

Without some precautionary measures, high-speed problems can creep into otherwise conventional digital designs. If a circuit is experiencing intermittent failures,

or if it encounters errors at voltage and temperature extremes, chances are there are some hidden signal integrity problems. These can affect time-to-market, product reliability, EMI compliance, and more.

Why is Signal Integrity a Problem?

Let's look at some of the specific causes of signal degradation in today's digital designs. Why are these problems so much more prevalent today than in years past?

The answer is speed. In the "slow old days", maintaining acceptable digital signal integrity meant paying attention to details like clock distribution, signal path design, noise margins, loading effects, transmission line effects, bus termination, decoupling and power distribution. All of these rules still apply, but…

Bus cycle times are up to a thousand times faster than they were 20 years ago! Transactions that once took microseconds are now measured in nanoseconds. To achieve this improvement, edge speeds too have accelerated: they are up to 100 times faster than those of two decades ago.

This is all well and good; however, certain physical realities have kept circuit board technology from keeping up the pace. The propagation time of inter-chip buses has remained almost unchanged over the decades. Geometries have shrunk, certainly, but there is still a need to provide circuit board real estate for IC devices, connectors, passive components, and of course, the bus traces themselves. This real estate adds up to distance, and distance means time—the enemy of speed.

It's important to remember that the edge speed—rise time—of a digital signal can carry much higher frequency components than its repetition rate might imply. For this reason, some designers deliberately seek IC devices with relatively "slow" rise times.

The lumped circuit model has always been the basis of most calculations used to predict signal behavior in a circuit. But when edge speeds are more than four to six times faster than the signal path delay, the simple lumped model no longer applies.

Circuit board traces just six inches long become transmission lines when driven with signals exhibiting edge rates below four to six nanoseconds, irrespective of the cycle rate. In effect, new signal paths are created. These intangible connections aren't on the schematics, but nevertheless provide a means for signals to influence one another in unpredictable ways.

At the same time, the intended signal paths don't work the way they are supposed to. Ground planes and power planes, like the signal traces described above, become inductive and act like transmission lines; power supply decoupling is far less effective.

EMI goes up as faster edge speeds produce shorter wavelengths relative to the bus length. Crosstalk increases.

In addition, fast edge speeds require generally higher currents to produce them. Higher currents tend to cause ground bounce, especially on wide buses in which many signals switch at once. Moreover, higher current increases the amount of radiated magnetic energy and with it, crosstalk.

Viewing the Analog Origins of Digital Signals

What do all these characteristics have in common? They are classic analog phenomena. To solve signal integrity problems, digital designers need to step into the analog domain. And to take that step, they need tools that can show them how digital and analog signals interact.

Digital errors often have their roots in analog signal integrity problems. To track down the cause of the digital fault, it's often necessary to turn to an oscilloscope, which can display waveform details, edges and noise; can detect and display transients; and can help you precisely measure timing relationships such as setup and hold times.

Understanding each of the systems within your oscilloscope and how to apply them will contribute to the effective application of the oscilloscope to tackle your specific measurement challenge.

Questions:

1) Why is ***camera*** mentioned in this article?

2) Can you tell something about EMI compliance?

3) What's the meaning of ***crosstalk*** in this article ?

4) Is the signal's rise time proportional to the processor's running clock rate?

5) Do you know how to track down the cause of a digital fault?

Passage 3 Virtual Instruments

What Is Virtual Instrumentation?

The rapid adoption of the PC in the last 20 years catalyzed a revolution in instrumentation for test, measurement, and automation. One major development resulting from the ubiquity of the PC is the concept of virtual instrumentation, which offers several benefits to engineers and scientists who require increased productivity,

accuracy, and performance.

A virtual instrument consists of an industry-standard computer or workstation equipped with powerful application software, cost-effective hardware such as plug-in boards, and driver software, which together perform the functions of traditional instruments.

Virtual instruments represent a fundamental shift from traditional hardware-centered instrumentation systems to software-centered systems that exploit the computing power, productivity, display, and connectivity capabilities of popular desktop computers and workstations.

Although the PC and integrated circuit technology have experienced significant advances in the last two decades, it is software that truly provides the leverage to build on this powerful hardware foundation to create virtual instruments, providing better ways to innovate and significantly reduce cost. With virtual instruments, engineers and scientists build measurement and automation systems that suit their needs exactly (user-defined) instead of being limited by traditional fixed-function instruments (vendor-defined).

Virtual Instruments versus Traditional Instruments

Stand-alone traditional instruments such as oscilloscopes and waveform generators are very powerful, expensive, and designed to perform one or more specific tasks defined by the vendor. However, the user generally cannot extend or customize them. The knobs and buttons on the instrument, the built-in circuitry, and the functions available to the user, are all specific to the nature of the instrument. In addition, special technology and costly components must be developed to build these instruments, making them very expensive and slow to adapt.

Virtual instruments, by virtue of being PC-based, inherently take advantage of the benefits from the latest technology incorporated into off-the-shelf PCs. These advances in technology and performance, which are quickly closing the gap between stand-alone instruments and PCs, include powerful processors such as the Pentium 4 and operating systems and technologies such as Microsoft Windows XP, .NET, and Apple Mac OS X.

In addition to incorporating powerful features, these platforms also offer easy access to powerful tools such as the Internet. Traditional instruments also frequently lack portability, whereas virtual instruments running on notebooks automatically incorporate their portable nature.

Engineers and scientists whose needs, applications, and requirements change very

quickly, need flexibility to create their own solutions. You can adapt a virtual instrument to your particular needs without having to replace the entire device because of the application software installed on the PC and the wide range of available plug-in hardware.

Flexibility

Except for the specialized components and circuitry found in traditional instruments, the general architecture of stand-alone instruments is very similar to that of a PC-based virtual instrument. Both require one or more microprocessors, communication ports (for example, serial and GPIB), and display capabilities, as well as data acquisition modules. What makes one different from the other is their flexibility and the fact that you can modify and adapt the instrument to your particular needs. A traditional instrument might contain an integrated circuit to perform a particular set of data processing functions; in a virtual instrument, these functions would be performed by software running on the PC processor. You can extend the set of functions easily, limited only by the power of the software used.

Lower Cost

By employing virtual instrumentation solutions, you can lower capital costs, system development costs, and system maintenance costs, while improving time to market and the quality of your own products.

Plug-In and Networked Hardware

There is a wide variety of available hardware that you can either plug into the computer or access through a network. These devices offer a wide range of data acquisition capabilities at a significantly lower cost than that of dedicated devices. As integrated circuit technology advances, and off-the-shelf components become cheaper and more powerful, so do the boards that use them. With these advances in technology come an increase in data acquisition rates, measurement accuracy, precision, and better signal isolation. Depending on the particular application, the hardware you choose might include analog input or output, digital input or output, counters, timers, filters, simultaneous sampling, and waveform generation capabilities. The wide gamut of boards and hardware could include any one of these features or a combination of them.

Software in Virtual Instrumentation

Software is the most important component of a virtual instrument. With the right software tool, engineers and scientists can efficiently create their own applications, by designing and integrating the routines that a particular process requires. They can also

create an appropriate user interface that best suits the purpose of the application and those who will interact with it. They can define how and when the application acquires data from the device, how it processes, manipulates and stores the data, and how the results are presented to the user.

With powerful software, you can build intelligence and decision-making capabilities into the instrument so that it adapts when measured signals change inadvertently or when more or less processing power is required.

An important advantage that software provides is modularity. When dealing with a large project, engineers and scientists generally approach the task by breaking it down into functional solvable units. These subtasks are more manageable and easier to test, given the reduced dependencies that might cause unexpected behavior. You can design a virtual instrument to solve each of these subtasks, and then join them into a complete system to solve the larger task. The ease with which you can accomplish this division of tasks depends greatly on the underlying architecture of the software.

Distributed Applications

A virtual instrument is not limited or confined to a stand-alone PC. In fact, with recent developments in networking technologies and the Internet, it is more common for instruments to use the power of connectivity for the purpose of task sharing. Typical examples include supercomputers, distributed monitoring and control devices, as well as data or result visualization from multiple locations.

Questions:

1) What factor brought forth the concept of virtual instrumentation?
2) What components does a typical virtual instrument have?
3) How can virtual instruments be user-defined?
4) What role does software play in virtual instrumentation?
5) What does the term ***distributed applications*** mean in this article?

参考译文

第 1 课　超大规模集成电路技术

晶体管是电子技术发展史上的一项关键发明，也是人类历史上的一项重要发明。1948 年，贝尔实验室(Bell Lab)发明了晶体管。简言之，晶体管是一种根据输入电量大小传导可变电量的器件。换言之，晶体管是一种数字开关。晶体管和真空管不同，它是“固态的”(solid state)。所谓“固态的”是指当晶体管改变状态时，其物理形式不发生变化。在晶体管中，不存在可以移动的部分。

和真空管相比，晶体管存在着巨大优势：晶体管的尺寸小得多，切换速度快得多，生产成本低得多；晶体管的性能更加可靠，耗能也较少。晶体管的发明带来了计算机工业的蓬勃发展。

最初的晶体管是分立器件。和其他分立器件一样，人们可以把晶体管单独置于电路当中。时至今日，一些特殊用途的晶体管还是分立器件形式的。集成电路的发明为现代处理器的诞生创造了条件。集成电路就是用单片材料制造的、内部互连的(无外部连线)一组晶体管。集成电路(IC)又叫芯片(chip)。

制造集成电路需要使用一类特殊材料。多数材料要么对电流绝缘(如空气、玻璃、木头)，要么很容易传导电流(如金属、水溶液)，但也有一些材料只能传导少量电流、或只在特定条件下传导电流——这种材料被称作“半导体”。硅是最常用的半导体材料。

通过精心的化学合成和设计，运用多种材料处理技术可以在硅层上直接构造极小的晶体管。这种晶体管体积小、速度快、性能可靠，耗能也少。1959 年，德州仪器公司(Texas Instruments)发明了第一片集成电路。它包含 6 只晶体管。

集成电路发明后不久，人们就意识到了“缩微化”(miniaturization)和集成大量晶体管到单片集成电路的巨大好处。为了实现更复杂的功能，就需要更多的晶体管(数字开关)。在提高硬件速度、控制功耗的同时，集成大量晶体管的关键就是“缩微化”技术。

“大规模集成”(LSI)是指将先前多个分立器件构成的电路集成化。LSI 器件包含几百个晶体管。早期计算机的构造就是在电路板上把多片 LSI 集成电路互连起来。

在大规模集成电路发明后，集成技术的水平随着时间的推移而不断提高。芯片变得越来越小，速度也越来越快，价格也更为低廉。在集成电路技术既有成果的基础上，工程师掌握了将更多逻辑集成到单片集成电路中的技术。这就是“超大规模集成”(VLSI)技术。VLSI 集成电路可包含几百万只晶体管。

早期处理器的功能是由多种逻辑芯片实现的。英特尔(Intel)公司率先将这些部件集成到单个芯片中。这就是第一片微处理器——英特尔公司于 1971 年推出的 4004。今天

的处理器已经非常先进了，却都是 4004 这个 4 位 CPU 的后代。

第 2 课　存储器件

存储器件的制造可以采用机械技术、磁技术、光技术、生物技术和电子技术。软盘、硬盘和铁电随机存储器属于磁存储器件。光存储器件有只读光盘、可写光盘。在计算机设备中，广泛使用的是电子存储器件。因为电子存储器件是当前速度最快的存储器件。在速度不太重要的应用中，常使用磁技术和光技术。

今天，所有电子存储器既可以是独立的集成电路形式、独立的模块形式，也可以作为集成电路的一部分。下表概括了几种电子存储器。（表略）

触发器（Flip Flop）

触发器是一种存储“0”或“1”的双态电路。由于结构简单，所以触发器速度极快。触发器是数字电路和集成电路的基础构件。由于电源电压去掉后，触发器的原有状态就失去了，因此触发器是“易失的”（volatile）。

寄存器（Register）

寄存器是一组并行触发器。寄存器的典型数据宽度为 8 位、16 位、32 位或者 64 位。寄存器常用于保存数据、地址指针等。和触发器一样，寄存器也是“易失的”，而且速度很快。

静态随机存取存储器（SRAM）

SRAM 是可寻址的触发器阵列。该阵列可配置成 1 位、4 位、8 位等数据格式。和触发器一样，SRAM 结构简单、存取速度快、具有“易失的”特点。在微控器电路板当中（芯片内部或外部）能发现 SRAM，因为在这些应用中所需的存储量不大，而且也不值得为了使用 DRAM 而去构建额外的接口电路。此外，因其存取速度快，SRAM 也用作高速缓存。

SRAM 的速度等级很多：从高速缓存的几个纳秒到低速应用的 200 ns。双极性 SRAM 和 MOS SRAM 现在都有。CMOS 技术的优势在于密度最高、功耗最低。高速缓存可以使用 BiCMOS 技术构建；BiCMOS 是一种混合技术，它使用双极性晶体管作为附加的驱动。采用“射极耦合逻辑”（ECL）双极性技术的 SRAM 具有最快的速度。由于这种技术的功耗高，所以存储器容量受到限制。

“内容寻址存储器”（CAM）是一种特殊的 SRAM 存储器。在这种技术中，构成存储

器的触发器阵列中的每一行都和一个数据比较器相连。访问存储器的方式不是向其提供地址，而是向其提供数据。所有的数据比较器将同时检查其对应的寄存器是否保存着和该数据相同的数据。CAM 将数据对应的行地址输出。该技术的主要应用是实现快速查找表。在网络路由器中，经常使用快速查找表。

动态随机存取存储器(DRAM)

“动态”是指数据没有保存在触发器中，而是保存在存储单元中。由于泄漏的存在，所以存储单元中保存的数据必须定期“更新”(即读出并重新写入)。更新时间间隔通常为 4～64ms。存储单元只需 1 个电容器和 1 个晶体管，而连接在阵列中的触发器则需 6 个晶体管。现代 DRAM 都使用沟道电容存储技术——晶体管置于电容之上，而使芯片尺寸最小化。因此，DRAM 技术的“比特成本”要低于 SRAM 技术。当所需存储量很大时，“数据更新需要额外电路”的缺点很容易就被较低的“比特价格”弥补了。

和 SRAM 存储器一样，DRAM 存储器也是由存储单元构成的。两者的主要区别之一在于寻址技术不同。对于 SRAM 存储器，需要为存储器提供地址，而存储器芯片输出存储单元中的数据；或者在输入端接收数据，并将其写入存储单元中去。对 DRAM 存储器而言，这种简单的存取方法是不可行的；由于动态的特点，读出一行数据而不再次将其写入会破坏该行内所有数据。

只读存储器(ROM)

ROM 也叫“掩模 ROM”或“掩模编程 ROM”。这是因为在制造 ROM 时，就需要通过将存储单元置“0”或置“1”来进行编程。通常，“0”或“1” 就是铝线的“有”和“无”。在制造芯片的最后一道工序中，铝层图案是由一块掩模平板决定的。所以，该类器件常被称作掩模 ROM。

“量产价格最低”是 ROM 的优点。对于某些应用来说，ROM 还具备另外一个优点——一旦芯片制造出来后，其中的数据就不能再改了，也就不需要进一步编程、测试了。但是，假如数据或代码非改不可的话，那么采用 ROM 就是“大错特错”。剩余的 ROM 芯片只好丢进垃圾桶，还得重新制造芯片。

电可擦除可编程只读存储器(EEPROM)

EEPROM 芯片和 EPROM 一样可编程，但编程采用的方法是“电擦除”。这种技术不需要紫外线信号源。EEPROM 支持“按字节擦除”。

第 3 课　微处理器

微处理器就是在单芯片上制造的、完整的运算引擎。第一片微处理器是 1971 年出现的 Intel 4004。4004 的功能不强，只能实现加法和减法，一次只能处理 4 比特数据。4004 引人注目之处在于全部功能集成在一片芯片上。在 4004 出现之前，工程师要用一组芯片或一组分立器件构建计算机。4004 的出现推动了便携式电子计算器的发展。

Intel 8080 是第一片用于家用计算机的微处理器。8080 是 1974 年出现的单片 8 位计算机。第一片获得商业成功的微处理器是 1979 年出现的 Intel 8088，它用于 IBM PC 中。PC 市场从 8088，80286，80386，80486 发展到奔腾Ⅰ、奔腾Ⅱ、奔腾Ⅲ和奔腾 4。这一系列微处理器都是由英特尔公司制造的，都是在 8088 基础上改进设计得来的。奔腾 4 可以执行任何一段曾在 8088 上运行的代码；不过，运行速度要快 5000 倍左右。下表展示了多年来英特尔公司生产的各种处理器之间的差别。（表略）

从上表可以看出：时钟速度和 MIPS 之间存在着某种联系。最高时钟速度是由生产工艺、片内延迟两个因素决定的。片内晶体管数量和 MIPS 之间也存在着某种关系。例如：时钟频率为 5 MHz 的 8088 的运行速率只有 0.033 MIPS（大约每 15 个时钟周期执行一条指令），而现代处理器的运行速度通常为每个时钟周期运行 2 条指令。运行速度的提高和片内的晶体管数量有直接关系。

微处理器的内部结构

微处理器执行一组机器指令。这些指令告诉微处理器去做什么。利用这些指令，微处理器可以完成以下三项基本任务：

1. 微处理器使用“算术逻辑单元”（ALU）来完成加、减、乘、除等数学运算。现代微处理器包含完整的浮点处理器，可以对大量浮点数据进行极其复杂的运算。

2. 微处理器可将数据从存储器的一个位置搬移到另一个位置。

3. 微处理器可以做出判断，并根据这些判断跳转到新的一组指令。

微处理器可以完成非常复杂的工作，但上述三项是最基本的。图 3.1 展示了一个能够完成上述三项工作的最简微处理器。

该微处理器有一套地址总线（向存储器发送地址）、一套数据总线（向存储器发送数据或者接收存储器数据）、一条读信号线 RD、一条写信号线 WR（用于通知存储器是从寻址地址读取数据、还是写入数据）、一条时钟信号线（为处理器安排时序的时钟脉冲）和一条复位信号线（将程序计数器置零，并重新开始执行）。这里假定数据总线和地址总线的宽度都是 8 位。

这个简易微处理器由如下组件构成：

1. 寄存器 A、寄存器 B 和寄存器 C：是由触发器构成的简易锁存器。

2. 地址锁存器：和寄存器 A，B，C 一样。

3. 程序计数器：一种具备“加一”功能和“置零”功能的锁存器。

4. 算术逻辑单元：可以简单到只是一个 8 位加法器，也可以是能够完成 8 位加、减、乘、除的单元（此处我们假定为后者）。

5. 测试寄存器：一种保存 ALU 比较结果的专用锁存器。通常，ALU 能够将两个数进行比较，并判断出二者是否相等或者哪个更大。测试寄存器也可以保存加法运算最后一步的进位位。这些数值保存在触发器当中，指令译码器利用这些数值做出判决。

6. 图中标有“3-State”的方框是三态缓冲器。它可以传送逻辑 1，逻辑 0，或者和输出断开。三态缓冲器允许在一条信号线上连接多个输出信号，但只有一个信号输出。

7. 指令寄存器和指令译码器负责控制所有其他组件。

尽管图中没有显示出来，但需要从指令译码器引出完成如下功能的控制信号线：

1. 通知寄存器 A 锁定出现在数据总线上的数值。

2. 通知寄存器 B 锁定出现在数据总线上的数值。

3. 通知寄存器 C 锁定出现在数据总线上的数值。

4. 通知程序计数器锁定出现在数据总线上的数值。

5. 通知地址寄存器锁定出现在数据总线上的数值。

6. 通知指令寄存器锁定出现在数据总线上的数值。

7. 通知程序计数器加一。

8. 通知程序计数器复位置零。

9. 激活六个三态缓冲器之一（六根独立的信号线）。

10. 通知 ALU 需要完成的操作。

11. 通知测试寄存器锁定 ALU 的测试位。

12. 激活 RD 信号线。

13. 激活 WR 信号线。

指令译码器的数据位不仅来自指令寄存器，而且来自测试寄存器和时钟信号线。

随机存取存储器和只读存储器

数据总线、地址总线、读/写信号线都连接到 ROM 上或者连接到 RAM 上（通常二者都有）。在这个微处理器例子中，有一套 8 位地址总线和一套 8 位数据总线。这意味着微处理器可寻址 256 字节的存储器，一次可以读/写 8 位数据。假定该微处理器有 128 字节（地址从 0 开始）的 ROM 和 128 字节（地址从 128 开始）的 RAM。

ROM 是只读存储器。ROM 芯片是用一组预设字节编程得到的。地址总线告知 ROM 芯片要将哪个字节取出并置于数据总线上。当 RD 信号线改变状态时，ROM 芯片

将选中的字节输出到数据总线上。

RAM 是随机存取存储器。RAM 中包含着以字节为单位的信息，微处理器能够依据 RD/WR 信号哪个有效来决定字节的读/写。当前 RAM 芯片的一个问题是：掉电后，所有保存在 RAM 上的内容全部丢失。这就是计算机需要 ROM 的原因。

顺便提一下，几乎所有计算机都有一定数量的 ROM（可以建造一种简单的不含 RAM 的计算机——许多微控器在片内集成了一定数量的 RAM——但是一般不可能建造出一种不含 ROM 的计算机）。在 PC 中，ROM 被称作"基本输入/输出系统"（BIOS）。当启动后，计算机就执行在 BIOS 中找到的指令。这些 BIOS 指令完成对机内硬件的测试，然后从硬盘中读取引导扇区。引导扇区也是一个小程序，BIOS 将其从硬盘中读出来之后，这个小程序就存储在 RAM 中。之后，微处理器开始从 RAM 执行引导扇区的指令。这个程序将告知微处理器从硬盘其他位置读取信息到 RAM 中，然后微处理器执行相应的指令等。这就是微处理器装载和执行整个操作系统的过程。

微处理器指令

这个简易微处理器的指令集也不小。指令集是以"比特组合"（bit pattern）方式实现的；当装载到指令寄存器的时候，每条指令都有各自的含义。由于人们不善于记忆"比特组合"，因此定义了一组短字来代表不同的比特组合。短字的集合就叫做处理器汇编语言。汇编器可以很容易将这些短字翻译成相应的比特组合，汇编器的输出被放置到存储器中以便微处理器执行。假如使用 C 语言进行编程，那么编译器会将 C 代码翻译为汇编语言。

这些指令在 ROM 中是什么样的呢？每一条汇编语言指令都用一个二进制数代表，这个二进制数称作"操作码"。指令译码器需要将操作码转换为驱动微处理器内部各组件的一组信号。以 ADD 指令为例，我们看一下具体要做什么。

在第一个时钟周期内，需要装载指令。指令译码器需要进行以下动作：

1. 激活程序计数器的三态缓冲器。
2. 激活 RD 信号线。
3. 激活三态缓冲器中的数据。
4. 将 ADD 指令锁存至指令寄存器。

在第二个时钟周期内，对 ADD 指令进行译码。微处理器仅仅需要完成：

1. 设置操作使 ALU 完成加法。
2. 将 ALU 运算结果锁存到 C 寄存器。

在第三个时钟周期内，程序计数器计数加一（在理论上，该操作可以并入第二个时钟周期）。

每条指令都可以这样分解为一组前后相继的操作，使微处理器的各个组件按照正确顺序工作。有些指令（如 ADD）可能需要 2～3 个时钟周期，有些指令可能需要 5～6 个时

钟周期。

微处理器性能

晶体管数量对于微处理器的性能有很大影响。如前所述，像8088这样的处理器执行一条典型指令需要15个时钟周期。由于设计乘法器的缘故，在8088上完成一次16位乘法需要大约80个时钟周期。晶体管越多，具备单周期乘法能力的乘法器就会越多。

有了更多的晶体管，就可以使用流水线技术。在流水线结构中，指令的执行是重叠的。尽管执行每条指令可能需要5个周期，却可以在不同阶段同时执行5条指令。这样看上去好像每个周期都能完成一条指令。

许多现代处理器有多个指令译码器，而每个指令译码器都有各自的流水线。这样，就可以实现“多指令流”——在一个周期内完成多条指令。“多指令流”技术实现起来相当复杂，需要使用大量晶体管。

目前，处理器的设计趋势是全32位ALU、内置快速浮点处理器和多指令流流水线。还有一个趋势是采用特殊指令，以便高效执行某些特定操作。此外，在处理器芯片上附加“硬件虚拟存储器”和“L1高速缓存”也是微处理器发展趋势之一。所有上述趋势都需要增加晶体管才能够实现；因此，今天已经出现了集成度高达几百万晶体管的微处理器。这些处理器可以在一秒内执行约10亿条指令。

第4课　运算放大器

那是在1934年，哈利·布莱克要从位于纽约市的家出发，乘坐火车或渡船去新泽西州的贝尔实验室上班。在上班途中，哈利能够放松下来，思考一些概念上的问题。哈利需要解决一个很棘手的问题：电话线在用于长途传输时需要放大器，而性能不可靠的放大器限制了电话业务的扩展。首先，放大器增益的容差性能很差；但是，通过使用调节器，这个问题很快就得到解决。第二，即使放大器在工厂里调节正确了，但在现场工作时增益还是漂移得很厉害，以至音量太低或输入语音发生畸变。

为了制造出稳定的放大器，哈利已经进行了多次尝试；但是，温度变化和电话线上出现的供电电压极限使得增益漂移无法控制。无源器件的漂移特性比有源器件好得多；假如能使得放大器增益仅由无源器件决定的话，那么这个问题就会解决。在乘渡船上下班的途中，哈利构思出了一个新颖的、解决放大器问题的办法，并在途中将它记录下来。

哈利的解决方案是：首先，设计一个增益比实际需求要高的放大器；然后，将放大器的一部分输出信号反馈到输入端，使电路增益（这里的电路是由放大器和反馈器件组成的）由反馈电路决定，而不是由放大器增益决定。这样的话，电路增益就取决于无源反馈器

件，而不是有源放大器。这个方案被称为“负反馈”，它是现代运算放大器的基本工作原理。哈利在乘坐渡船途中记录了第一个有意加入反馈的电路。此前，也一定有人无意中使用过反馈电路，但设计者却忽视了这种效应！

管理者和放大器设计者可能会发出痛苦的抱怨。他们也许会说：“获得 30 kHz 的“增益带宽积”（GBW）就够难的了，现在这个傻瓜要我设计 GBW 为 30 MHz 的放大器，而他还是想得到一个 GBW 为 30 kHz 的电路”。然而，时间已经证明哈利是正确的；不过，有一个小问题，哈利没有详细讨论。那就是振荡问题。在环路闭合的时候，开环增益很大的电路有时会发生振荡。很多人都研究过这种不稳定现象；不过，直到 20 世纪 40 年代，人们对它才有了清晰的认识。可是，解决稳定性问题需要长时间单调、复杂的计算。好多年过去了，没有人能将解法简化或者使之易于理解。

1945 年，伯德（Bode）提出了一种用于分析反馈系统稳定性的图形化方法。在这之前，反馈分析是用乘法和除法完成的。因此，计算传输函数是一项耗时、费力的工作。要知道：20 世纪 70 年代，工程师们才有了计算器和计算机。伯德采用的是一种对数方法——将分析反馈系统稳定性的数学过程转换为既简单又好理解的图形化分析。反馈系统设计依然很复杂，却再也不是一项只为少数电气工程师所掌握的技术了。哪个电气工程师都能用“伯德法”去判定反馈电路的稳定性，反馈技术也越来越多地应用于机器设计中。而对电子反馈设计真正迫切的需求是在计算机和传感器成为成熟技术后才产生的。

第一台实时计算机是模拟计算机。这台计算机使用预先编程的数学公式和输入数据来计算控制动作。编程是对一系列电路进行“硬连线”，这些电路对输入数据进行数学操作。“硬连线”的局限性导致模拟计算机无法普及。运算放大器是模拟计算机的核心部件。配置好的运算放大器可以对输入数据进行各种数学运算（如乘法、加法、减法、除法、积分和微分）。随着人们逐步了解、喜欢运算放大器，其名称就简化为大家熟知的“op-amp”（运放）。运放采用了具有很大开环增益的放大器；当电路形成闭合环路时，放大器就会执行由外部无源元件控制的数学运算。这种放大器是真空管制作的，而且需要大电压电源供电，所以体积就很大。不过，人们还是容忍了运放的“大体积”和“大电压”——因为它是模拟计算机的核心部件。早期运放多是为模拟计算机设计的，但人们很快就发现运放还有其他用途；此后，在物理实验室里很容易见到运放了。

那时，在大学和大公司的实验室里，能够看到通用模拟计算机——因为计算机对研究工作是至关重要的。同时，在实验中也需要对传感器信号进行调理；在信号调理方面，运放也找到了用武之地。随着信号调理应用范围的拓展，其对运放需求的增长超过了模拟计算机。在模拟计算机落伍于数字计算机之后，运放并未受到冷落，因为它在通用模拟应用中的重要性。最终，数字计算机替代了模拟计算机；而随着测量应用的增加，运放的需求量也增长了。

在晶体管出现之前，第一个用于信号调理的运放是用真空管构建的，因此它的体积很大。20 世纪 50 年代，低电压工作的小型真空管使运放的体积缩小到一块砖的大小，因此

运放模块有了一个“砖块”的绰号。真空管和组件的体积不断缩小，直至运放缩至一只八脚真空管的大小。在 60 年代，晶体管实现了商业开发，这进一步将运放的体积减至几立方英寸。早期运放大多是为特定应用制造的，所以它们不一定通用；早期运放是为某种特定应用服务的，而每个生产商制定的技术指标和封装都各不相同；所以，早期运放几乎都不存在另外的货源。

在 20 世纪 50 年代末、60 年代初，集成电路开发出来了。但直到 20 世纪 60 年代中期，仙童公司才发布了 μA709。μA709 是首片取得商业成功的集成运放。虽然 μA709 有自身的问题，但任何一位称职的模拟工程师都会使用它，在多种不同模拟应用中都可以使用它。μA709 的主要缺陷在于稳定性；它需要外部补偿，需要称职的模拟工程师去使用它。μA709 也很敏感；一旦不利条件出现，它就容易自毁。在 μA709 之后，出现了内部补偿的 μA741；当在数据手册要求的条件下工作时，μA741 是不需要外部补偿的。从那以后，新型运放就不断出现；如今，运放的功能和可靠性已经提高到了“人人都能在模拟应用中使用”的程度。

今天，集成运放依然存在。最新一代的运放可以覆盖 5 kHz～1 GHz GBW 的频谱范围。供电电压范围从可靠运行的 0.9 V 到最大绝对指标的 1000 V。输入电流和输入偏置电压低到用户在验证指标时遇到难题。运放已经成为名副其实的“通用”模拟集成电路，因为它可以完成所有模拟任务——线路驱动器、比较器（一位模数转换）、放大器、电平转换器、振荡器、滤波器、信号调理器、执行部件驱动器、电流源和电压源等。而设计人员所面临的则是如何迅速选定恰当的电路（运放的组合）以及如何计算出无源元件值来产生期望的传输函数。

运放是如此重要，它将继续作为模拟设计的关键部件。每一代新型电子设备都在硅片上集成了更多的功能，并将更多的模拟电路置于集成电路之中。随着数字应用的增加，模拟应用也会增加——因为大量的数据接口应用是在现实世界中，而现实世界是一个模拟的世界。因此，每一代新型电子设备都需要新型模拟电路，也就需要新一代运放去满足相应的需求。未来，模拟设计和运放设计依然是一项必需的技能。

第 5 课　低通滤波器

一阶滤波器

从数学公式上讲，积分器（图 5.1a）是最简单的滤波器；它是构成大多数现代滤波器的基本模块。我们怎么从直观上理解积分器呢？假设在输入端加一个直流信号（即频率为 0），那么在输出端将会出现一个线性斜坡信号——其幅度一直增至电源电压。如果不

考虑电源电压对输出信号的限制，积分器在零频率上的响应将是无穷大；这意味着它在零频率点上存在一个极点（在使传输函数为无穷大的频率点上存在一个极点）。

我们也知道，积分器的增益随着频率的增加而减小；在很高频率上的输出电压实际上就是 0。增益和频率成反比，因此在对数坐标系中是一条斜率为-1的直线（图 5.1b）。

传递函数很容易推导出来（公式略）；其中，s 是复频率变量 $\sigma+\mathrm{j}\omega$，ω_0 等于 $1/RC$。假如把 s 当作频率的话，这个公式印证了我们有关“增益和频率成反比”的直观判断。

简单低通 RC 滤波器（图 5.2a）是一种比积分器稍复杂的滤波器。其传输函数如下：（公式略）。当 s 等于 ω_0 时，函数简化为 1；当 s 趋近于无穷大时，函数简化为 0；因此，这是一个低通滤波器。当 s 等于 $-\omega_0$ 时，分母为 0 且函数值为无穷大，这意味着在复频率平面有一个极点。图 5.2b 画出了传递函数幅度和复频率 s 的关系，s 的实部（σ）指向读者，而 s 的正虚部（$\mathrm{j}\omega$）朝右。$-\omega_0$ 处的极点非常明显。为了突出函数的形状，幅度是用对数值显示的。在无穷大频率处，积分器和 RC 低通滤波器的频率响应都趋近于 0，也就是说在 s 等于无穷大处存在一个零点。这个零点环绕着复平面。

s 变量复函数和电路的实际频率响应是如何联系起来的呢？在分析电路对交变信号的响应时，我们用 $\mathrm{j}\omega L$ 来表示电感的阻抗，用 $1/\mathrm{j}\omega C$ 来表示电容的阻抗。在使用拉普拉斯变换分析瞬态响应时，我们用 sL 和 $1/sC$ 分别表示电感的阻抗和电容的阻抗。这种相似性是显而易见的。实际上，交流分析中的 $\mathrm{j}\omega$ 就是 s 的虚部；正如上面提到的那样，s 是由一个实部 σ 和一个虚部 $\mathrm{j}\omega$ 组成的。

假如将上面任何一个公式中的 s 用 $\mathrm{j}\omega$ 替代，就可以得到电路的角频率响应。在图 5.2b中，σ 等于 0，s 等于 $\mathrm{j}\omega$ 且沿正 j 坐标轴 。因此，沿着该坐标轴的函数值就是滤波器的频率响应。我们沿着 $\mathrm{j}\omega$ 坐标轴将函数切割，并且在沿正坐标轴的函数值处加以粗线，从而突出了 RC 低通滤波器的频率响应曲线。因为频率是对数表示的缘故，所以曲线和人们熟悉的伯德图看上去不一样（图 5.2c）。

复频率的虚部（$\mathrm{j}\omega$）有助于描述电路对交流信号的响应，而实部（σ）有助于描述电路的瞬态响应。从图 5.2b 中，我们可以看出 RC 低通滤波器和积分器之间的区别。RC 低通滤波器的瞬态响应更加稳定，因为其极点位于复平面的左半部。即对于阶跃输入，滤波器的响应是衰减指数形式的。对于 RC 低通滤波器而言，极点沿$-\sigma$坐标轴离原点越远意味着 ω_0 越大，时间常数越短，瞬态响应越快。与此相反，极点离 j 坐标轴越近，瞬态响应越慢。

到目前为止，我们已经阐述了几个简单电路的数学传递函数与其复频率平面上的极、零点之间的关系。从这些函数中，我们可以推导出电路的频率响应（继而可以得到伯德图）及其瞬态响应。在积分器和 RC 滤波器各自的传递函数中，分母中都只有一个 s，所以它们只有一个极点。也就是说，它们都属于一阶滤波器。

从图 5.1b 中，我们也可以看出：一阶滤波器的频率选择性不是很好。为了得到更接近需求的滤波器，我们必须增加滤波器的阶数。从现在开始，我们将使用 $f(s)$ 表示传递

函数，而不再使用 V_{OUT}/V_{IN}。

二阶低通滤波器

二阶滤波器的分母中出现了 s^2，所以在复平面就有两个极点。使用电感和电容的无源电路或者包含电阻、电容和放大器的有源电路都具有这样的响应。例如，考虑图 5.3a 中的无源 RLC 滤波器。其传输函数具有如下形式：(公式略)。当定义 $\omega_0{}^2 = 1/LC$ 和 $Q = \omega_0 L/R$ 之后，该式变为：(公式略)，此处的 ω_0 是滤波器的特征频率，而 Q 是品质因数(R 越低，Q 越高)。

极点出现在使分母为零的 s 值处，即当 $s^2 + s\,\omega_0/Q + \omega_0^2 = 0$ 的时候。只要记得 $ax^2 + bx + c = 0$ 的根为(公式略)，我们就可以解出这个方程。

在这个方程中，$a=1, b=\omega_0/Q$，而 $c=\omega_0^2$。$(b^2-4ac) = \omega_0^2(1/Q^2-4)$，所以如果 Q 小于 0.5，那么两个根都是实数且位于负实轴上。而电路行为和两个一阶 RC 滤波器级联非常相似。这种情况意义不太大，所以我们只考虑 $Q > 0.5$ 的情况。这意味着(b^2-4ac)小于零，而方程的根都是复数。

实部为$-b^2a$，等于$-\omega_0/2Q$，这部分对于两个根都是一样的。两根的虚部互为相反数。计算根在复平面的位置，我们就会发现它们位于距零点 x_0 处(图 5.3b)。

改变 ω_0 会改变极点距离原点的位置。降低 Q 值会将两个极点靠近，而增加 Q 值会将极点沿着半圆弧拉开并趋近 j 轴。当 $Q = 0.5$ 时，两极点在负实轴上的 $-\omega_0$ 处汇合。在这种情况下，对应的电路就等效于两个一阶滤波器的级联。

现在，让我们考察一下二阶函数的频率响应及其随 Q 变化的情况。图 5.4a 展示了在由复频面和垂直幅度向量构成的三维空间中绘制的函数曲线面。图中 $Q = 0.707$，可以立即看出这是一个低通滤波器的响应。

增大 Q 值的效果是将极点沿着圆弧路径移向 $\mathrm{j}\omega$ 轴。图 5.4b 展示的是 $Q = 2$ 的情形。由于极点靠近 $\mathrm{j}\omega$ 轴，其对频率响应的影响就更大，这导致在通带高频端出现了一个峰值。

在滤波器的瞬态响应上也有影响。因为极点的负实部越小，输入阶跃函数就会引起滤波器输出振铃。Q 值越小，振铃越少，因为阻尼更大。另一方面，假如 Q 为无限，极点抵达 $\mathrm{j}\omega$ 轴就会引起在 $\omega=\omega_0$ 处的频率响应为无限大(不稳定及持续振荡)。在图 5.3a 的 RLC 电路中，这种情形是不可能出现的，除非 $R=0$。然而，对于包含放大器的滤波器而言，这种情形是有可能出现的。在设计过程中，必须要考虑到这一点。

二阶滤波器提供了 ω_0 和 Q 这两个变量，这就允许我们将极点放置在复平面上任何需要的地方。然而，这两个极点必须是一对共轭对。二者实部相同、而虚部符号相反。极点放置的灵活性是一种强大的手段，它使二阶滤波器成为众多开关电容滤波器中的有用部件之一。随着频率趋近于无穷，二阶低通滤波器的传递函数和一阶低通滤波器传递函数一样趋近于零。然而，二阶低通滤波器传递函数的下降速度是一阶低通滤波器传递函数

下降速度的两倍;这是分母中出现 s^2 因子的缘故。因此,在无穷大处就存在二重零点。

第 6 课　模数转换器(ADC)

模数转换器(ADC)的种类和冰激凌的口味一样多,挑选 ADC 需要像挑选冰激凌一样细心。

最常见、最易理解的 ADC 是闪式 ADC(见图 6.1)。闪式 ADC 能够实现高速转换,因而能提供很高的采样率。不过,基本形式的闪式 ADC 的功耗很大。

闪式 ADC 将输入信号同时置于一列(2^N-1 个)比较器之前,比较器的参考电压由一组电阻设定,参考电压精确对应于转换器能代表的全部采样电压。比较器的输出(0 或 1)被编码成 N 位码字以代表输入采样信号电平。这是最简单、最直观和最快速的 ADC 实现方案。对于多数实际应用而言,当位数较多时(如大于 14 位),所需的电阻数量就太大了。与一些速度稍慢、特殊的模数转换方案相比,闪式 ADC 的功耗明显要高。

图 6.2 展示了典型闪式 ADC 的输出频谱。这张图是通过对 ADC 转换后的纯正弦样本进行“快速傅里叶变换”(FFT)得到的。可以看出:频谱不是简单地由输入正弦分量组成,还有一些分布在测量带宽内的其他成分。这些成分主要来自无法避免的“量化误差”(噪声)。这是因为转换器要使用有限个可用样值(由 ADC 位数决定)来代表模拟输入电平。

对于满量程正弦波,有效“信噪比”(SNR)是转换器精度的函数;当满足某些条件时,它可由如下公式决定(公式略)。其中,N 是转换器的位数。

产生图 6.2 中曲线的 ADC 具备 12 位转换精度,其理论 SNR 值为 74 dB。从该曲线中,我们可以看出:正弦分量电平和单个噪声分量电平之差接近 104 dB。这两个数值之间存在差异的原因是 SNR 公式使用了全部噪声,其中包括构成 FFT 的全部噪声分量之和。指出这一点的原因是:通过提高采样率,可以提高转换器的有效精度(见下文)。

过采样实现“处理增益”

在某个给定的音频应用中,假定我们需要转换过程中的信噪比至少为 70 dB。上述公式表明:需要最少 12 位的转换精度(前提是满幅正弦波输入和理想转换器),测量中包含($0\sim0.5f_s$)Hz 带宽中的全部噪声。如果采用 8 倍带宽的采样率(图 6.3),我们可以看到实际的音频信号仅占用了基带(即 $0\sim0.5f_s$)的 1/4,而噪声在带宽内均匀分布。假如现在准备对采样信号进行数字滤波,我们可以去掉约 3/4 的噪声,从而将信噪比提高 4 dB 或 6 dB。这种有效 SNR 的提高称作“处理增益”,是通过对输入信号进行“过采样”(相对于 2 倍带宽原则而言的)来实现的。

可以很容易推导出计算处理增益的简易公式:处理增益(dB)=10log(采样率/2×信

号带宽)。在该公式中,我们假定:(1)为了准确符合期望的输入信号带宽,我们使用数字滤波器对样本带宽进行限制;(2)噪声是均匀分布的。

这样的话,假如我们使用128倍过采样设计(常见于迷你盘录音设备和PC机声卡),那么就可以获得18 dB的"信号/量化噪声比"的提高,从而将转换器的有效精度从12位提高至15位。顺便说一个例子,我们可以考虑使用该方法提高数字移动电话中数据转换器的性能。一个GSM蜂窝信道的带宽为200 kHz。目前,我们可以得到采样率为80 MSPS、精度为14位的高速模数转换器;这种ADC在整个($0\sim f_s/2$)范围内提供75 dB信噪比。其处理增益如下:处理增益(dB) = 10 log(80 000 000/2× 200 000)= 26 dB。因此,GSM采样信号具备很高的信噪比:75 + 26 = 101 dB。

"西格玛—德尔塔"转换器

"处理增益"的概念恰好可以引领我们进入"西格玛—德尔塔"(Σ—Δ)转换器这个话题。由于充分利用了"处理增益"的概念,Σ—Δ转换器获得了高性能、低成本和低功耗的优势;Σ—Δ转换器只需简单的模拟接口(需要很少的或者不需要抗混叠滤波器),所以非常适合于音频应用。图6.4展示的是基本Σ—Δ转换器框图。

这个转换器实质上是一个后跟数字滤波和抽取的高过采样1位ADC(即比较器),数字滤波和抽取用来实现处理增益。由于增加了噪声整形电路,该转换器的有效性能获得了极大提高。噪声整形电路将原本在$0\sim f_s/2$频带内均匀分布的噪声最大限度地从有用频带中去除(图6.5)。

对于典型的128倍过采样系统,仅"处理增益"一项就可将有效精度提高3位(即相当于一个4位转换器)。而"噪声整形"技术能将有效精度进一步大幅提高;对于某些用于音频的Σ—Δ转换器而言,现在已经可以达到24位精度了。现代转换器使用了比图6.5复杂得多的噪声整形处理方法。尽管图6.5只是简单的一阶Σ—Δ设计,但采用"处理增益"和"噪声整形"的基本原则却是一致的。

目前,在实际中使用的模数转换方法有几十种之多。常见的方法有逐次逼近法、多通道法、插值法、子区法和逐位处理法。每种方法都具备一些其他方法不具备的潜在优势。幸运的是,在选择ADC器件时,用户一般不需要理解其工作原理。只要用心研究一下数据手册上提供的性能指标,就可以做出最佳选择。

第7课 开关电源

除了电池供电的电子产品外,新型电子产品都需要将115 V或230 V交流电源转换为直流电源为电路供电。而电能转换效率正在成为业界以及整个社会关注的重点。

开关电源不仅提供了较高的转换效率，而且为设计者提供了更大的灵活性。半导体技术、磁器件和无源器件技术的发展使开关电源在当今功率转换的舞台上成为日益流行的选择。

线性电源与开关电源

在历史上，线性稳压器曾经是产生稳压输出电压的主要方法。通过对级联功率通过器件导电性进行线性控制以响应负载变化，线性稳压器将输入高电压降为输出低电压。这种方法导致在负载电流流经的通过单元两端出现一个大电压。

这种功率损耗使线性稳压电源的效率只有 30％到 50％。这意味着每向负载输送 1 W的功率，就会有至少 1 W 的功率以热能形式消耗了。对于那些功率超过 10 W 的小型应用而言，散热装置的成本就导致使用线性稳压电源不合算。

开关电源中的功率器件工作在“全开”和“截断”状态。这样，要么在大电流流经功率器件时，导通电压很低；要么在大电压时，没有电流通过器件。因此，电源内部消耗的功率就很少。开关电源的平均效率为 70％～90％，而且和输入电压无关。

集成度的提高带动了开关电源成本的下降，这使开关电源在输出功率超过 10 W 及多输出应用中成为具有吸引力的选择。

基本转换器

前向模式转换器基础

最基本的前向转换器是如图 7.1 所示的降压转换器。其工作过程可视为两个不同的时间周期，它们分别出现在串联功率开关处于“接通”和“断开”的状态。当功率开关接通时，输入电压连接到电感的输入端。电感的输出电压就是转换器的输出电压，整流器处于反向偏置。在这个周期，由于在电感两端存在恒定电压源，所以电感电流开始按照如下公式线性增加：(公式略)。

在“接通”周期，能量以磁通的形式存储在电感的铁芯材料中。存储的能量足以满足负载在下一个“断开”周期的需求。

下一个周期就是功率开关的“断开”周期。当功率开关断开时，电感输入电压被捕获二极管钳位在地电位以下一个二极管电压降。电流开始流过捕获二极管以维持负载环路电流。存储在电感中的能量被移走。这段时间内的电感电流可表示为：(公式略)。当功率开关再次接通时，这个周期就结束了。

电压调节是通过改变功率开关的“接通－断开”的占空比来实现的。下面的关系式近似描述这个工作过程：(关系式略)，式中的 ∂ 是占空比。

该转换器可以输出 kW 级功率，但却有一个严重缺点——假如功率开关非正常短路，输入电源就会直接连接到负载电路，这通常会造成灾难性后果。为了避免出现这种情况，在输出端跨置了一个“短路器”(crowbar)。短路器是一个闭塞的“可控硅整流器”(SCR)。当输出

进入“过电压”(overvoltage)状态，就会激发这个整流器。降压转换器仅用于板级电源管理。

增压模式转换器基础

最基本的增压模式转换器是升压稳压器(见图 7.2)。其工作过程也可以分为两个不同的周期，分别对应着功率开关的“导通”和“断开”状态。当功率开关导通时，输入电压源接到电感两端。电流从零开始按照如下公式增加：(公式略)。能量被存储在铁芯材料中。

每个周期存储的能量乘以工作频率的积必须大于负载的功率需求，即满足下式(关系式略)。

然后，功率开关断开，电感电压超过输入电压并被整流器钳位在输出电压上。电流开始线性下降直至铁芯中的能量全部耗尽。如图 7.3 所示，其电流波形由下式决定(公式略)。升压变换器也仅用于板级电源管理。

拓扑结构

拓扑结构是指功率器件和磁元件的布局。每种拓扑结构在特定应用中都有各自的优点。有若干因素决定着一种拓扑结构对于某个特定应用的适用性：

(1)该拓扑结构的输入、输出之间是否实现了电气隔离？

(2)有多少输入电压加在电感或者变压器两端？

(3)流经功率半导体器件的峰值电流为多少？

(4)是否需要多个电压输出？

(5)功率半导体器件两端出现的电压有多大？

设计者面临的第一个选择是：是否采用输入和输出之间的变压器隔离。非隔离开关电源的典型应用是为具备绝缘隔板的系统提供板级电压调理。非隔离拓扑结构还可用于这种情况：当出现故障时，输入电源不会连接到易损负载电路。对于所有其他情况，应该使用变压器隔离。和变压器隔离相关的是对多输出电压的需求。变压器为开关电源增加附加输出提供了一种易于实现的方法。自己制造系统电源的公司倾向于在尽可能多的电源中采用变压器隔离，因为这种隔离避免了故障出现时将产生的连锁反应。

第 8 课　时钟信号源

时钟器件类型

今天的时钟器件多种多样，下面阐述其中几种。

晶体

晶体是一种基本的压电石英晶体。它本身是不能产生时钟信号的，必须和时钟振荡

器连接在一起才能得到时钟波形。晶体有两种:串联谐振晶体(可视作高品质因数的串联LC电路)和并联谐振晶体(可视作高品质因数的并联LC电路)。串联谐振晶体在谐振频率点的阻抗最小,而并联谐振晶体在谐振频率点的阻抗最大。

晶体振荡器

晶体振荡器是一种用晶体反馈元件组成的振荡器,其他振荡器采用有源器件或无源器件作为反馈元件。晶体振荡器的输出频率最为精确和稳定。晶体振荡器是多数高速数字系统时钟源的首选。

补偿振荡器

随着温度和电压的变化,晶体振荡器的输出频率也会发生变化。在需要高稳定度时钟的应用中,人们通常使用补偿振荡器。补偿振荡器能够调整由电压和温度引起的频率变化。“温度补偿振荡器”(TCXO)包含了用于补偿温度变化的电路,从而防止频率的变化。恒温控制振荡器将晶体放置在一个温控恒温箱中,这样保持晶体工作在一个精确温度之下。双恒温箱振荡器有两个恒温箱,晶体在内层恒温箱中,而控制电路和内层恒温箱又包含在外层恒温箱中。双恒温箱振荡器比恒温控制振荡器的温度稳定性更好。显然,随着温度稳定性的提高,振荡器的成本也提高了。

压控振荡器

“压控振荡器”(VCO)的输出受输入电压引脚的控制。在整个频率范围,控制电压和输出频率的关系是非线性的,但是在部分频率范围内是线性的。

频率合成器

通过使用一个或多个“锁相环”(PLL),频率合成器从一个或多个参考时钟源产生一个或多个不同的输出频率。参考频率通常是由连接到合成器上的晶体产生的。设计频率合成器的目的是用以替代系统中的多个振荡器,从而减少了电路板空间,降低了系统成本。图8.1展示了锁相环的框图。

PLL有两个输入,一个参考输入和一个反馈输入。PLL采用两种方法校正频率。“频率校正”先对参考输入和反馈输入间的大频差进行校正。“频率校正”类似于“粗调”;当VCO的频率低于参考频率的一半或者高于参考频率的二倍时,要进行“频率校正”。当VCO的频率在参考频率的一半和二倍之间时,要进行“相位校正”;“相位校正”是“微调”。

“相位/频率检测器”检测参考输入和反馈输入之间频率差和相位差,并依据反馈频率超前还是滞后于参考频率分别产生用于补偿的“Up”信号和“Down”信号。然后,这些控制信号通过“电荷泵”和“环路滤波器”产生控制VCO的控制电压。振荡器的频率取决于控制电压信号。VCO的稳态频率为$F_{\mathrm{vco}}=F_{\mathrm{ref}}(P/Q)$。PLL的输出频率可以表示为$F_{\mathrm{out}}=(F_{\mathrm{ref}}\cdot P)/(Q\cdot N)$。

频率合成器的采样率决定了为进行相位和频率校正而对输入信号采样的频率。其表达式为F_{ref}/Q。

基于 PLL 的频率合成器的“采集/锁定时间”是频率合成器在加电后(或在可编程输出频率发生改变之后)达到目标频率所用的时间。

基于 PLL 的频率合成器的“精度”是 P、Q 计数器的位数。精度决定了频率改变的增量。

基于 PLL 的频率合成器的“死区”是指无法被 PLL 校正的参考输入和反馈输入之间的最大相位差。

产生多个不相关频率的频率合成器需要使用多个 PLL。随着系统复杂性的提高和系统中多个时钟的使用,频率合成器应用得越来越普遍。“时钟信号产生器”和“频率合成器”这两个词可以互换使用。

时钟缓冲器

时钟缓冲器是一种输出波形直接跟随输入波形的器件。输入波形通过该器件并被输出缓冲器重新驱动。因此,该类器件存在传输延迟。此外,由于在每个输入、输出通道间存在传输延迟的差别,输出端将出现“相位抖动”。

时钟器件参数

时钟抖动

“抖动”被定义为时钟输出的状态转换位置偏离了理想位置。这种偏离可能超前于理想位置,也可能滞后于理想位置。因此,“抖动”表示为 ± ns。“抖动”可分为“周期间抖动”、“周期内抖动”和“长期抖动”。

“周期间抖动”是相邻周期的周期长度差。这种抖动最难测量,通常需要时序间隔分析仪。

“周期内抖动”也称为“短期抖动”。它是指在相继时钟边沿的范围内时钟输出状态转换位置偏离理想位置。注意:“短期抖动”测量的是时钟上升沿对理想位置的偏离,并用时间单位或频率单位表示。

“长期抖动”是指在“很多”周期范围内时钟输出状态转换位置对理想位置的偏离。“很多”的具体数字取决于应用和频率。对于个人计算机主板和图形应用,这个数字通常是指 10 到 20 ms。对于其他应用,这个数字可能不同。

“抖动”的成因主要有如下四项:电源噪声、合成器内部锁相环、晶体或其他谐振器件的随机热噪声和晶体振动产生的随机机械噪声。

时钟抖动影响几乎所有的高速同步系统。常见的受抖动影响的应用有:个人计算机主板、图形卡和通信设备。

相位偏移

“相位偏移”是指应该同时到达的两路信号在抵达时间上的差异。“相位偏移”由两部分组成:驱动器件的“输出相位偏移”和由于布线导致的“电路板设计相位偏移”。

时钟驱动器相位偏移(又称“内部相位偏移”)是由时钟驱动器引起的相位偏移量。时

钟驱动器件有缓冲器件和基于锁相环的器件两类。缓冲器件的相位偏移出现在输出端，因为输入信号通过器件的传输延迟各不相同。这种不同主要归因于输出负载的不同。对于基于锁相环的时钟器件而言，其相位偏移非常小，因为它可以通过调整来补偿输出负载的不同。

电路板设计相位偏移(又称"外部相位偏移")是由下列电路板布线问题造成的：

• 迹线长度：信号通过一条迹线所用的传输时间是由印制电路板材料、迹线长度、迹线宽度和容性负载决定的。迹线长度的不同导致信号传输时间的不同，从而引起相位偏移。

• 门限电压差别：接收器件的门限电压也会导致相位偏移。例如，假设某个接收器件的门限电压为 1.2 V，而另外一个接收器件的门限电压为 1.7 V，输入信号的上升时间为 1 V/ns，那么这两个器件的状态切换时刻就会相差 500 ps，这就是相位偏移。

• 容性负载：迹线容性负载的不同会导致负载端时钟上升时间的不同。这将影响时钟边沿在何时超过输入门限电平，从而导致相位偏移。

• 传输线端接：由于当今时钟驱动器的边沿速率极高，长度超过 4 英寸的迹线就应视作"传输线"。如果没有正确进行端接，迹线就会出现像"电压反射"这样的传输线效应，从而导致相位偏移。

为什么"相位偏移"如此重要？在高速系统中，时钟的"相位偏移"是构成系统时序极限的重要部分。对于 15 ns 的时钟周期而言，1 ns 的"相位偏移"就是很明显的。如果在时序设计的时候没有考虑到"相位偏移"问题，那么系统很可能无法可靠地工作。

测量两路输出信号间"相位偏移"的最简单方法是用双通道示波器显示两路信号的波形，并测量两者上升沿的时间差，就可以得到相位偏移。

容差/精度

这个参数度量的是器件工作频率和标称频率(一般是指在常温下)的接近程度。例如，某个器件的标称输出频率为 25 MHz，而其在常温下输出频率的"长期"(由用户定义)平均值为 25.001 MHz，那么该器件的精确度为＋40 ppm(ppm 为百万分之一)。频率"容差"要受到晶体生产和校准工艺中的"精度控制能力"的影响或控制。

稳定性

"稳定性"这个参数通常和晶体、振荡器相关。稳定性是指在常温范围内工作频率和常温标称频率之间的差别(ppm)。该参数由频率偏差的最大值和最小值给出(%或ppm)。稳定性为什么重要？假如在设计时没有考虑到稳定性，可能会导致该设计在整个温度范围内处于极限工作状态。

老化

"老化"被定义为内部晶体或振荡器随时间的变化而引起的频率上的系统误差。"老化"通常用"ppm/年"来表示。如不单独表述的话，"老化"可以和"稳定性"指标合在一起。该参数通常和晶体振荡器有关。新晶体的老化速度要高于旧晶体。

摆动

电压或频率的变化率称为“摆动”。通常,“摆动”是在数字信号的上升沿和下降沿进行测量的。然而,在生产商提供的产品手册中,更常见的是“上升时间”和“下降时间”,而不是“摆动”。近来,随着低功耗器件的出现,“摆动”用来定义频率变化率。

占空比

“占空比”是输出信号高电平时间和整个周期时间之比。该参数用百分比表示。理想的“占空比”是50%,而多数时钟生产商指定“占空比”为40% ~ 60%。在既使用时钟上升沿也使用时钟下降沿的系统中,“占空比”是很重要的。

TTL器件和CMOS器件均有“占空比”。对于TTL器件,电压在0 ~ 3 V之间摆动,“高电平时间”在1.5 V测量。对于CMOS器件,电压在0 ~ V_{dd}之间摆动,“高电平时间”在$V_{dd}/2$测量。

第9课　互连部件

互连部件(Interconnect)用于连接器件、电路板和模块。在设计阶段,就必须考虑到对互连器件的需求。

互连部件必须满足下列大多数或者全部性能要求:信号完整性、功率损耗、电气特性(接触电阻,电感和电容,电压和功率额定值,屏蔽,滤波)、机械特性(接触点数量,间隔和力,冲击和振动,尺寸,焊接技术)和环境问题(清洗材料,热性能,防腐保护)。

“信号完整性”是指输入信号的质量损失。在理想的情况下,互连部件输出端信号和输入端信号的所有特性完全相同。在实际中,会发生信号质量的下降。用户必须明确可接受的质量下降程度,或者必须定义输出端信号最小可接受电平。能够反映信号完整性的测试包括“电压驻波比”(VSWR)、频率响应、信号边沿上升时间和电流。

高速连接器问题

在今天的设计中,时钟速率超过100 MHz,上升时间通常为1 ns或更低;设计者不能无视互连部件对逻辑设计的影响。对电路的时序特性和噪声特性,互连部件都有显著的影响。时钟速率和上升时间的加快增加了容性耦合效应和感性耦合效应,这使得串扰问题更加严重。这也意味着在数据定时、读取前,反射信号衰退的时间变短了,这就要求减小“未端接系统”的最大导线长度。这一切都意味着一个重大挑战:确保高速脉冲经过互连路径(从器件到印制电路板、通过印制电路板到底板、并进而延伸到任何正在使用的网络连接)时的信号完整性。

因为出现了反射,所以在通过互连路径传送高速信号脉冲时,短促的上升时间就会破

坏信号的完整性。上升时间越短，信号质量下降的风险越大。理想的情况是：连接器拥有合适的端接特性，没有造成信号质量的下降。

串扰

除了信号质量下降之外，“串扰”(crosstalk)的增加是另一个对信号完整性的潜在影响。串扰是由接触点间的感性和容性耦合引起的。它能够将一条导线(或接触点)中的开关信号噪声感应到另一条导线(或接触点)。串扰可以表示为信号电压摆幅的百分比或者dB数；0 dB代表100%的串扰，−20 dB代表10%的串扰，−40 dB代表1%的串扰。

串扰分成“后向串扰”和“前向串扰”两类。“后向串扰”也叫“近端串扰”，它在连接器的驱动端进行测量，代表了容性耦合和感性耦合之和。“前向串扰”也叫“远端串扰”，它在连接器的接收端进行测量，代表了容性耦合和感性耦合之差。

串扰还可分为“单线串扰”和“多线串扰”。“单线串扰”是一条驱动线感应电压至一条静止线的结果。“多线串扰”是两条或两条以上驱动线感应电压至一条静止线。大多数系统要同时使用多条导线(多个接触点)，因此“多线串扰”是一个更为实用的数据。

为了全面理解连接器的串扰指标，用户需要获取以下信息：

1. 串扰指标是实际测量结果，还是仿真结果？
2. 串扰测量时使用的信号“上升时间”是多少？是用何种数字技术产生的(TTL和CMOS“上升时间”的典型定义为幅度的10%到90%，而ECL“上升时间”的典型定义为幅度的20%到80%)？
3. 连接器上的信号模式是什么？例如，地线相对于可测量信号线的位置。
4. 信号状态切换影响了多少条线？
5. 源阻抗和负载阻抗是多少？
6. 串扰指标是指“前向串扰”、“后向串扰”，还是全部串扰？
7. 串扰指标是指“单线串扰”，还是“多线串扰”？
8. 如果是“多线串扰”，有几条驱动线？

“在信号接触点间分布接地点”是减小串扰的常用方法。当然，这种方法的缺点是降低了信号接触点密度。

互连的层次

戈兰尼兹定义了互连的六个层次：

第一层：芯片焊盘到封装引线的连接(如导线接头)

第二层：元器件到电路板的连接(如PLCC插座)

第三层：电路板间的连接(如用于主板和子板连接的边缘卡)

第四层：部件互连(如扁平电缆)

第五层：部件和输入输出的互连(如BNC连接器)

第六层：系统互连（如以太网连接器）

每一层次的互连都可以有多种实现方式。例如，第四层电路板间互连可以采用边缘卡连接器、扁平电缆或导线。

第 10 课　无线移动电话系统(Ⅰ)

无线移动电话分为两大类：无绳电话和移动电话。无绳电话由基站和室内手持设备组成。无绳电话不支持网络传输。移动电话的发展经历了三个不同的时期，各有不同的技术——模拟语音、数字语音、数字语音和数据（互联网、电子邮件等）。

第一套移动系统是由美国 AT&T 公司设计的，美国“联邦通信委员会”（FCC）委托该公司将其应用于全美国。因此，美国各地使用了统一（模拟）系统——在加州购买的移动电话到了纽约州也可以使用。在欧洲，移动电话系统的情形与美国正好相反——欧洲各国都设计了自己的系统。这种做法导致了移动电话在欧洲的失败。

从这次失败中，欧洲汲取了教训。当数字移动通信技术到来时，政府运营的邮电公司联合推出了统一的系统标准（GSM）；在欧洲境内任何地点，任何一部欧洲移动电话都能够使用。而当时，美国决定政府不应参与标准化事务，而将数字移动通信的发展交由市场决定。该决定导致美国现有两大（加上一个较小的）互不兼容的数字移动通信系统。

尽管美国最早引领了移动通信技术的应用，但是现在欧洲移动电话的拥有率和使用率远远超过了美国。“拥有欧洲全境统一的系统”是原因之一，但还有其他原因。美国和欧洲另一个不同之处在于“电话号码”这件小事上。在美国，移动电话号码和固定电话号码是不分的。这样一来，用户看不出某个号码（如(212)234－5678）是固定电话号码还是移动电话号码。为了避免人们在使用电话时的顾虑，电话公司决定移动电话用户只为来电付费。这样做的结果是：由于担心仅接电话就会积累大额账单，许多人在购买移动电话时很犹豫。在欧洲，移动电话都有一个特殊的区位号（类似 800 号码和 900 号码），这样就很容易识别移动电话号码。这样，“主叫付费”的惯例也可以应用于欧洲的移动电话（费用分开的国际电话除外）。

还有一个对采用移动电话的产生巨大影响的因素是预付费移动电话在欧洲的广泛使用（在某些地区，高达 75％的移动电话是预付费的）。在许多商店里，人们都可以像买收音机一样地买到这种电话——“付完钱就可直接拿走”。你既可以预存话费（如 20 或 50 欧元），也可以在余额降为零时对其充值。因此，欧洲的儿童和许多小孩儿都拥有移动电话（通常是预付费电话）；这样，孩子的父母就可以知道孩子在哪里，而孩子们也不会花掉大量话费。如果仅仅偶尔使用移动电话，那么使用移动电话实质上是免费的，因为不存在月租费和来电费。

第一代移动电话:模拟话音

在 20 世纪前叶,移动电话偶尔用于航海和军事通信。1946 年,第一套车载电话系统在圣路伊斯架设起来。该系统位于很高的建筑物顶上,使用一台发射机和一个同时用于接收、发送的信道。为了进行通话,用户必须按下一个按钮启动发射机、禁止接收机。20 世纪 50 年代末,有几个城市开始架设这种系统(称为“按键—通话”系统)。“民用无线电”(CB)设备、出租车和警车经常使用这种技术。

20 世纪 60 年代,“改进移动电话系统”(IMTS)架设起来。IMTS 系统也在山顶上使用大功率(200 W)发射机;不过,现在拥有两个频率——一个用于发射、另一个用于接收。由于移动电话的输入信号和输出信号占用不同的信道,移动用户无法听到对方(不同于出租车上的“按键—通话”系统)。IMTS 系统提供 150 MHz 到 450 MHz 之间的 23 个信道。由于信道数量很少,用户往往得等待很长时间才能听到拨号音。此外,位于山顶的发射机功率很大,临近系统必须相距几千米以避免互相干扰。总之,由于容量有限,IMTS 系统并不实用。

先进移动电话系统(AMPS)

随着“先进移动电话系统”(AMPS)于 1982 年在贝尔实验室发明及其在美国的架设,一切都发生了改变。AMPS 系统也用于英国,并被称为 TACS;AMPS 在日本被称作 MCS－L1。为了实现向下兼容,AMPS 系统的许多基本特性都由其后来者 D－AMPS 继承下来。

在移动电话系统中,地理区域分成一个个蜂窝状的小区。这就是“移动电话”也称作“蜂窝电话”的原因。在 AMPS 系统中,蜂窝小区的典型大小为 10～20 km;在数字系统中,蜂窝小区要小一些。每个小区使用一组特定的频率,这组频率在其临近的任何小区都不能使用。使蜂窝系统具备很大容量的关键在于使用了较小的蜂窝小区和在附近(但不是临近的)小区重复使用频率。在方圆 100 km 内,IMTS 系统的每个频率上只能拥有一路通话;在同样区域内,AMPS 系统拥有 100 个 10 km 大小的蜂窝小区,而且每个频率拥有 10～15 路通话。这样,蜂窝设计将系统容量至少提高了一个数量级,蜂窝越小、提高量越大。此外,蜂窝越小意味着需要的功率越小,发射机和手持设备会更小、更便宜。手持电话输出功率为 0.6 W,车载发射机功率为 3 W,这是 FCC 允许的功率最大值。

图 10.1(a)展示了“频率重用”的概念。通常,蜂窝小区都接近圆形,但用六边形进行建模更加容易。在图 10.1(a)中,蜂窝小区的大小都相同。每 7 个小区分成一组。每个字母代表一组频率。需要注意:对于每组频率,存在约两个小区宽的缓冲地带;为了实现更好的隔离以降低干扰,在这些缓冲地带,频率是不能重用的。

在用户数量达到导致系统过载程度的地区,为了实现更多的频率重用,需要降低功率并将“过载”蜂窝小区分为更小的“微蜂窝”单元(图 10.1(b))。在体育赛事、摇滚音乐会及其他在几个小时内有大量移动用户云集的场所,电话公司有时会使用卫星链接的便携

发射塔创建临时“微蜂窝”。

基站位于蜂窝小区的中心，小区内的所有呼叫都传送到基站。基站由计算机和发射机/接收机（与天线相连）组成。在较小系统中，所有基站都和一个称作“移动电话交换局”（MTSO）或“移动电话交换中心”（MSC）的设备相连。在更大的系统中，需要几个MTSO。所有MTSO都和二级MTSO相连，……，以此类推。MTSO实质上就是电话系统中的端局；实际上，它至少和一个电话系统端局相连。MTSO使用“分组交换”网络和基站、其他MTSO及“公共交换电话网”（PSTN）进行通信。

第11课　无线移动电话系统(Ⅱ)

第二代移动电话：数字话音

第一代移动电话是模拟的，第二代移动电话是数字的。第一代移动电话没有世界通用标准，第二代移动电话也没有世界通用标准。目前，正在使用的有D-AMPS、GSM、CDMA和PDC这4个系统。PDC只用于日本，它基本上就是为和日本第一代模拟系统兼容而对D-AMPS做的修改。

数字先进移动电话系统(D-AMPS)

第二代AMPS系统是全数字的D-AMPS系统。国际标准IS-54以及后来的IS-136描述了D-AMPS系统。为使同一蜂窝小区中的一代、二代移动电话能够同时使用，D-AMPS系统进行了精心设计以便和AMPS共存。特别需要指出的是：在相同频率上，D-AMPS系统采用了和AMPS相同的30 kHz信道；这样一来，某个模拟信道的附近可能是数字信道。

当D-AMPS开始服务时，为了解决载荷增加的问题，就需要使用一个新频带。上行信道在1850～1910 MHz范围内，相应的下行信道在1930～1990 MHz范围内；上行信道和下行信道是成对出现的，这一点和AMPS系统完全相同。该频段的波长是16 cm，标准“1/4波长”天线仅4 cm长；这样一来，电话尺寸就减小了。然而，为了获得更大范围的可用信道，许多D-AMPS电话机既使用850 MHz频段，也使用1900 MHz频段。

在D-AMPS移动电话上，语音信号由麦克风拾取之后，先要对其进行数字化，然后要使用比“增量调制”和“预测编码”更为复杂的模型进行压缩。由于考虑到了人类发声系统的详尽特性，这种压缩算法能把标准PCM编码的56 kbps带宽压缩到8 kbps或更低。压缩过程是由“声码器”电路完成的。为了减少空中链路传送的数据量，压缩过程是在话机内完成的，而不是在基站或端局完成的。对固定电话来说，在话机内完成压缩不会带来任何好处——因为降低本地环路的通信量根本不会增加系统容量。

对于移动电话而言，在话机内完成“数字化”和“压缩”会带来巨大收益——利用“时分复用”技术，一个“频率对”可被 3 位 D-AMPS 用户共享。每个“频率对”支持的帧速率为 25 帧/秒(即每帧占用 40 ms)。而每帧又分为 6 个 6.67 ms 的“时隙”(time slot)最低频率对的情形如图 11.1(a)所示。

每帧拥有 3 位用户，这 3 位用户轮流使用上行链路和下行链路。例如，在图 11.1(a)“时隙 1”期间，“用户 1”可能在向基站发送数据，而“用户 3”在接收基站传来的数据。每个时隙的长度是 324 位。其中，64 位用于保护时间、同步和控制，其余 260 位供用户使用。在这 260 位当中，101 位用于噪声环境下空中链路所需的误码校正，因此最终只有 159 位用于“压缩语音”数据。由于每秒内包含 50 个时隙，所以“压缩语音”数据的可用带宽恰好低于 8 kbps——标准 PCM 带宽的 1/7。

如果使用更好的压缩算法，还可以把语音数据率降至 4 kbps；这样的话，一帧中就可以容纳 6 位用户了(见图 11.1(b))。在运营商看来，“3～6 位 D-AMPS 用户只需占用相当于 1 位 AMPS 用户所需的带宽”是巨大的成功，而且在很大程度上解释了“个人通信业务”(PCS，Personal Communicaiton Service)普及的原因。

全球移动通信系统(GSM)

D-AMPS 广泛应用于美国，其改进形式应用于日本。世界其他地区使用的是一种称作 GSM (全球移动通信系统)的系统，美国也开始有限规模地使用 GSM 系统了。乍一看，GSM 和 D-AMPS 很相似。两者都是蜂窝系统、都使用“频分复用”(一个频率用于发送，一个更高的频率用于接收；D-AMPS 系统的接收频率比发送频率高 80 MHz，而 GSM 系统的接收频率比发送频率高 55 MHz)、都采用“时分复用”方法将一个频率对分成多用户共享的几个时隙。不过，GSM 系统的信道比 D-AMPS 系统要宽很多(GSM 信道宽度为 200 kHz，D-AMPS 信道宽度为 30 kHz)，而增加的用户数相对要小(一个 GSM 信道的用户数为 8，一个 D-AMPS 信道的用户数为 3)。因此，GSM 系统为用户提供的数据率比 D-AMPS 系统高得多。

如图 11.2 所示，每个频带的宽度为 200 kHz。GSM 系统拥有 124 对单工信道。每个单工信道带宽为 200 kHz，并以时分复用的方式提供了 8 路独立连接。每个正在使用的基站都会分配到频率对上的一个时隙。从理论上讲，每个蜂窝小区可以支持 992 个信道；但为了避免和邻近蜂窝小区发生频率冲突，许多信道并不能使用。在图 11.2 中，8 个阴影时隙全部属于同一个连接，每个方向 4 个。在同一时隙中，不会同时进行发送和接收；这是因为 GSM 系统不能同时发送和接收射频信号，发送和接收之间的切换是需要时间的。假定分配到 890.4/935.4 MHz 频点上时隙 2 的移动站向基站发送数据的话，它会使用靠下的 4 个阴影时隙(以及时间上相继的时隙)，并且向每个时隙添加数据直至所有数据发送完毕。

图 11.2 所示的 TDM 时隙只是一个复杂组帧结构的局部。每个 TDM 时隙都有特定的结构，而若干组 TDM 时隙又构成了具有特定结构的“复帧”(multiframe)。该结构的简

化示意如图 11.3 所示。从这张图中,我们可以看到:每个 TDM 时隙包含一个 148 位的数据帧,它占用了 577 μs 的时间(包括每个时隙之后的 30 μs 保护时间)。数据帧的开头和结尾都是 3 个"0",这几位用来对帧进行描述。数据帧还包含两个 57 位的"信息字段"(information field),每个字段包含 1 位"控制位"(control bit)。"控制位"用来标明下一个信息字段是用于语音还是用于数据。信息字段之间有一个 26 位的"同步字段"(又称"训练字段"),接收机用它来同步发送方的帧边界。

一个数据帧占用的发送时间是 547 μs,而发射机只允许每隔 4.615 ms 发送一个数据帧。原因是每个发射机都要和其他 7 个发射机共享信道。每个信道的总数据率为 270.833 kbps,这个数据率要被 8 位用户等分。这样,每位用户得到的总数据率为 33.854 kbps,这是 D-AMPS 系统中单用户总数据率(50×324 bps=16.2 kbps)的两倍还多。

从图 11.3 中可以看出,8 个数据帧构成一个 TDM 帧,而 26 个 TDM 帧构成一个 120ms 的"复帧"。在这 26 个 TDM 帧当中,"时隙 12"用于控制,"时隙 25"保留使用,所以只有 24 个时隙用于传输用户数据。

码分多址(CDMA)

CDMA 和 AMPS、D-AMPS 及 GSM 完全不同。CDMA 没有将可用频带分割为几百个窄带信道,而是允许信号在整个频谱范围内传送。利用编码原理,CDMA 可以将多个同时传送的信号分离开来。在 CDMA 中,"相遇数据帧会造成数据混淆"的想法不存在了——CDMA 认为多路信号之间是线性相加的。

我们不妨打个比方:在机场候机厅内,三三两两的人正在攀谈。TDM 就像一个人站在屋子中间轮流和别人交谈;FDM 就像分散在各处的人群,每群人都同时进行交谈,但却互不干扰。CDMA 就像屋子中一直在说话的人,但是每对谈话者都使用不同的语言。讲法语的那一对想听到的语言是法语,而把其他语言当作噪声。因此,CDMA 的关键在于能够提取出希望得到的信号,而把其他信号当作随机噪声加以抑制。

CDMA 的"位时间"(bit time)被分成 m 个称作"码片"(chip)的、更短的时间间隔。每位包含的典型码片数为 64 或 128。为简单起见,下例采用的是 8 码片/位。

每个站分配到一个唯一的称为"码片序列"的 m 位码。要传送比特"1",就需要发送该站的码片序列;要传送比特"0",就需要发送该站码片序列的反码。发送其他序列是不允许的。因此,对于 $m = 8$ 的情形,假如 A 站分配的码片序列为 00011011,它通过发送 00011011 发送"1"比特,通过发送 11100100 发送"0"比特。

如果可用带宽提高至原来的 m 倍,发送信息量就可从"b 位/秒"提高到"mb 码片/秒"。这使得 CDMA 成为"扩频通信"方式之一(假定调制和编码技术没有发生变化)。假设 100 个站的可用带宽为 1 MHz,如果采用 FDM,每个站将拥有 10 kHz 带宽、数据率可达 10 kbps(假定信道利用率为 1 bit/Hz);如果采用 CDMA,每个站则可以利用 1 MHz 带宽,码片速率为 1 Mc/s。如果每个比特所含码片数小于 100 的话,那么每个 CDMA 站

的有效带宽要高于 FDM，信道分配的问题也解决了。

随着样本位数越来越多，无噪声“奈奎斯特”(Nyquist)信道具有任意大容量；同理，在理想的无噪声 CDMA 系统中，系统容量(即站点数)也为任意大。在实际应用中，因受到各种物理限制，系统容量明显减少了。首先，我们假定所有码片在时间上是同步的——而这在实际中是无法实现的。可以做到的是：发送方和接收方通过传送预先确定的码片序列来实现同步，该码片序列的长度足以使接收机能够锁定。所有其他的(未同步的)传送均被视为随机噪声。正如人们可以预期的那样，码片序列越长，在噪声环境中正确检测出信号的概率就越高。为了更加可靠，比特序列可以使用纠错编码。码片序列从不使用纠错编码。

我们的讨论隐含了这样的假设：在接收机看来，所有发射站的功率电平是一样的。无线通信是 CDMA 的典型应用领域。在无线通信系统中，有一个固定的基站和许多与基站距离随时变化的移动站。基站接收到的信号功率电平取决于它和发射机之间的距离。可以直观地推断出如下结论：移动站向基站发送信号的功率是接收到的基站信号功率电平的倒数。换言之，移动站从基站接收到一个弱信号要比接收一个强信号耗费更多的功率。基站也可以向移动站发出明确指令以使之提高或降低发射功率。

我们还假定接收机知道发送方是谁。理论上讲，如果提供了足够的计算能力，通过并行运行解码算法，接收机可以同时听到全部发送方。但实际的情况是“说起来容易、做起来难”。在这里，我们也没有谈到许多其他使问题更加复杂化的因素。但无论如何，CDMA 确实是一种非常巧妙的设计方案，并正在迅速地应用于无线移动通信之中。CDMA 的工作带宽为 1.25MHz(D-AMPS 的工作带宽为 30 kHz，GSM 的工作带宽为 200 kHz)；在这个带宽范围内，CDMA 支持的用户数要远远超过其他系统。在实际中，CDMA 用户的可用带宽至少和 GSM 一样出色，而且往往比 GSM 要好很多。

第 12 课　个人计算机系统

一提到“技术”这个词，多数人都会联想到计算机。其实，我们生活的方方面面都存在着和计算机有关的事物。诸如电视机等家用电器的内部就有微处理器，甚至汽车里也有一个计算机。不过，人们首先想到的还是“个人计算机”或者叫做“PC”。

PC 是一种以微处理器为中心的通用工具。它由存储器、硬盘、调制解调器等很多不同的部件共同构成。所谓“通用”是指使用 PC 可以做许多不同的事情。既可以用它录入文件，也可以用它发送电子邮件、浏览网页和玩游戏。

PC 内部

让我们看一看一台典型台式计算机有哪些主要部件。

• 中央处理单元(CPU):计算机系统的微处理器“大脑”叫做“中央处理单元”。CPU 监管着计算机所做的一切。

• 内存:用于保存数据的快速存储设备。内存速度快的原因是它直接和微处理器相连。计算机有几种类型的存储器:

1)随机存取存储器(RAM,用于暂存计算机正在使用的信息)。

2)只读存储器(ROM,一种永久性存储器,用于存储不变的重要数据)。

3)基本输入输出系统(BIOS,一种用于计算机开机时建立基本通信的 ROM)。

4)缓存(用于在极高速 RAM 中存储频繁使用的数据,直接和 CPU 相连)。

5)虚拟存储器(用于暂存数据和在需要时和 RAM 交换数据的硬盘空间)。

• 主板:所有其他内部组件都和主板连接。通常,CPU 和内存都位于主板上。其他组件可能直接位于主板上,也可能通过某种方式连接到主板上。例如,声卡可以置于主板上,也可以通过 PCI 总线连接到主板上。

• 电源:计算机用来调节电能的电气转换器。

• 硬盘:用于保存程序、文档等信息的大容量永久存储设备。

• 操作系统:为用户和计算机之间提供接口的基础软件。

• 集成驱动器电路(IDE)控制器:硬盘驱动器、只读光盘驱动器和软盘驱动器的主要接口。

• 周边设备互连(PCI)总线:这是一种最常见的将附加设备连接到计算机的方式,PCI 总线使用了一系列主板插槽和 PCI 卡相连。

• 小型计算机系统接口(SCSI):SCSI 是一种向计算机添加硬盘、扫描仪等附加设备的方法。

• 加速图形接口(AGP):AGP 是一种用于显卡和计算机进行接口的超高速连接。

• 声卡:计算机使用声卡进行音频信号的录放。方法是将模拟信号转换成数字信息,然后再将数字信息转换为模拟信号。

• 显卡:把来自计算机的图像数据转换为显示器可以显示的数据格式。

PC 连接

输入/输出设备(I/O)

不论计算机内部组件的功能多么强大,这些组件之间都需要互相作用。这种相互作用称为“输入/输出”(I/O)。PC 中最常见的几种 I/O 设备如下。

• 显示器:显示器是显示计算机信息的主要设备。

• 键盘:键盘是向计算机输入信息的主要设备。

• 鼠标器：鼠标器是浏览计算机、和计算机交互作用的主要设备。

• 可移动存储设备：有了可移动存储设备，向计算机添加信息变得非常容易，同时也便于将信息保存并携带到其他地方。

1）软盘：软盘是最常见的可移动存储设备。软盘价格极其便宜，存储信息非常容易。

2）只读光盘（CD-ROM）：CD-ROM 是商业软件最流行的发布方式。现在，许多系统提供可刻录光盘 CD-R 和可擦写光盘 CD-RW。

3）闪存：以一种称作“电可擦写只读存储器”（EEPROM）的“只读存储器”（ROM）为基础，闪存能够提供快速、永久的数据储存。CompactFlash 卡、SmartMedia 卡和 PCMCIA 卡都属于闪存。

4）数字视盘（DVD-ROM）：DVD-ROM 和 CD-ROM 类似，但能够保存的信息要多得多。

端口

• 并行口：常用于连接打印机。

• 串行口：典型的应用是连接外置调制解调器。

• 通用串行总线（USB）：USB 很快成为了最流行的外设连接方式，USB 端口功能强大、配置灵活、极易使用。

• FireWire（IEEE 1394）：FireWire 是一种非常流行的将数字视频设备（如便携式数码摄像机、数码相机等）和计算机进行连接的方法。

互联网/网络连接

• 调制解调器：这是和互联网连接的标准方法。

• 局域网（LAN）卡：许多计算机都使用局域网卡，尤其是用于以太办公网内计算机的互连。

• 线缆调制解调器：有些人在家中使用有线电视系统上网。

• 数字用户线（DSL）调制解调器：在标准电话线上高速工作的一种连接。

• 超高速数字用户线（VDSL）调制解调器：VDSL 是一种较新的改进 DSL，它需要具有光缆连接的电话线。

从开机到关机

在熟悉了 PC 的组成部件之后，现在让我们看一看计算机从开机到关机的整个过程。

1. 按下计算机和显示器的“开”（on）按钮。

2. 可以看到 BIOS 软件在运行“开机自检”（POST）程序。在许多机器中，BIOS 以文本方式显示存储器容量、硬盘类型等信息。在引导过程中，BIOS 完成了大量的工作，为计算机的运行做准备。

1）BIOS 判定显卡是否可以使用。多数视频卡自带一个小型的 BIOS，这个 BIOS

对卡上的存储器和处理器进行初始化。如果没有这么做的话，BIOS 通常可以从主板上的另一个 ROM 中装载显卡驱动信息。

2）通过查看存储器地址 0000:0472 处的值，BIOS 检查这是“冷启动”还是“重新启动”。如果这个值是 1234h，就表明这是“重新启动”；BIOS 就跳过余下的开机自检步骤。如果是任何其他数值，就被当作是一次“冷启动”。

3）如果是冷启动，那么 BIOS 通过对每一个存储器地址进行数据读/写来验证 RAM 的有效性。然后，BIOS 检查键盘和鼠标。BIOS 寻找 PCI 总线；如果找到了，就检查所有的 PCI 卡。在开机自检过程中，如果 BIOS 发现了什么错误，它会通过发出一连串“哔哔”声或在屏幕上显示一条文本消息来通知用户。这个时候出现的错误几乎都是硬件问题。

4）BIOS 查看在 CMOS 设置中作为引导设备的存储设备的顺序。“Boot”是“bootstrap”这个词的缩写形式，这个“bootstrap”就是用在“Lift yourself up by your bootstraps”这句老话中的那个“bootstrap”。“引导”是指启动操作系统的过程。BIOS 使用“引导程序”对设备依序进行初始化。

3.“引导程序”将操作系统装载到内存并使之开始运行。“引导程序”是通过建立起用于保存操作系统、用户信息和应用程序的存储器分区来完成这项工作的。之后，“引导程序”建立起用于在子系统内部、子系统之间、应用程序内部和应用程序之间进行通信的数据结构。最后，它将计算机的控制权交给操作系统。

一旦操作系统载入后，就需完成以下六大类任务：

- 处理器管理：将任务分解为易管理的若干部分，并在向 CPU 发送之前为其分配优先级。
- 内存管理：协调出入 RAM 的数据流，决定何时需要虚拟内存。
- 设备管理：提供每个连接到计算机上的设备和 CPU、应用程序之间接口。
- 存储管理：将数据永久保存在硬盘或其他形式存储设备的相应位置。
- 应用程序接口：提供软件程序和计算机之间的标准通信和数据交换。
- 用户接口：提供人机交互和通信的途径。

当你选择了“关机”(shut down)选项，操作系统就会关闭所有正在运行的程序。假如某个程序有尚未保存的信息，在程序关闭前会给用户一次保存的机会。操作系统将它当前的设置写入一个专门的配置文件中，这样下次启动时就会有相同的设置。假如计算机提供了对电源的软件控制，那么操作系统在完成自己的关机任务后将计算机完全关掉。如果不是这样，用户就得亲自将电源关闭。

第 13 课　现代数字设计概览

传统上，电子电路设计可分为模拟设计、数字设计两大类。这两类课程通常是分别教授的，电子工程师一般只精通一个领域。这两类设计还可以进行专业细分。例如，射频模拟电路设计、数字集成电路设计和混合信号电路设计（当数字和模拟两大领域相遇时）。此外，在嵌入式系统设计中，软件工程师也扮演着越来越重要的角色。

在消费类电子产品当中，数字电子技术更加重要。汽车拥有复杂的控制系统。许多家庭如今也拥有了个人电脑。曾被视为模拟产品的收音机、电视机和电话也都在数字化。数字“光盘”（CD）差不多完全替代了“密纹唱片”（LP）。这些变化的出现使得产品的生命周期缩短了。在不到一年的时间里，附近商店里的原有产品很可能就完全被新型数字电子产品替代了。

设计自动化

为了紧跟如此迅速的变化，电子产品必须要以极快的速度设计出来。模拟设计依然是专业化程度很高、收入颇丰的职业之一。而数字设计已经非常依赖“计算机辅助设计”（CAD）——又称“设计自动化”（DA）或“电子设计自动化”（EDA）。EDA 工具可以完成“综合”（synthesis）和“模拟”（simulation）这两类任务。“综合”是将设计指标转变为实际的设计实现，而“模拟”是对设计指标或详细的实现方案进行“演习”（exercise），以确保其正确运行。

用于综合和模拟的 EDA 工具要求设计者将想法转化成设计。这可以通过使用“图形化软件包”（graphical package）绘制设计原理图来实现。这种方法称为“原理图输入”。另一种方法是用文本形式来表示设计，这种方法和软件编程很相似。对数字电路硬件的文本描述可以采用编程语言（如 C 语言），也可以采用“硬件描述语言”（HDL）。在过去约 30 年中，设计出来了几种 HDL。Verilog 和 VHDL（VHSIC 硬件描述语言，VHSIC 代表超高速集成电路）是两种广泛应用的硬件描述语言。标准化 HDL 是很重要的，因为标准化语言可以被不同生产商提供的不同 CAD 工具使用。在 Verilog 和 VHDL 出现之前，每个 CAD 工具都有自己的 HDL。在不同 HDL 之间进行转换（例如，使用其他模拟器来验证某个综合工具的输出）既费时又费力。

逻辑门

门电路是构成数字电路的基本模块。门电路是由几个输入和（通常情况下）一个输出组成的电子器件。在通常情况下，输入和输出是“逻辑 0”和“逻辑 1”这两个状态之一。逻

辑值可以由电压(例如,0 V 代表“逻辑 0”,3.3 V 代表“逻辑 1”)或电流来表示。门电路利用全部输入实现某种逻辑运算以产生输出。当然,数字门电路说到底也是模拟器件。不过,为了简单起见,我们倾向于忽略其模拟本质。

我们可以买到一片如图 13.1 所示的集成电路,它包含 4 个完全一样的门电路(注意:有两个连接用于器件的正、负电源。一般情况下,这些连接不显示在逻辑原理图中)。将成百上千个这样的器件连接起来,可以构成一个数字系统——实际上,许多系统就是这样设计的。虽然每片集成电路的成本也许只有区区 10 美分,可用于印制电路板设计和系统组装的花销却是非常可观的——采用这种设计方法并不合算。

大规模生产的集成电路(从“触发器”一直到“微处理器”)集成了越来越多的复杂功能。日益增加的复杂度带来了灵活性的提高——微处理器通过编程可以完成几乎无限种任务。因此,某些数字系统的设计过程就是“选择标准元器件、将其连接起来”。不过,标准器件可能无法提供某种功能,这是避免不了的。设计者只好选用“分立门电路”(discrete gates)来实现这种功能,或者设计专用集成电路来实现这种功能。“设计专用集成电路”或许令人畏缩,但不要忘记——系统成本在很大程度上取决于器件互连成本、而非单个器件成本。

专用集成电路(ASIC)和现场可编程门阵列(FPGA)

高性能“全定制集成电路”(full－custom IC)设计是一项难度很大的工作。在全定制集成电路设计中,所有事情(甚至每个晶体管)都需要设计(尽管可以使用元件库)。不过,利用“门阵列”(gate arrays)构造“半定制集成电路”(half－custom IC)已有很多年。顾名思义,“门阵列”就是集成了逻辑门阵列的集成电路。所以,使用门阵列设计“专用集成电路”(ASIC)就是要定义阵列中的门电路之间的连接。从实现角度看,就是要设计一层或两层的金属互连线。集成电路的制造流程分为七个步骤(或更多),除去最后一步“金属化”(metallization)之外,其他步骤都可提前完成。由于门阵列可以大量生产,所以单个器件成本就比较低。

ASIC 通常是指全定制和半定制集成电路。“可编程逻辑”是另外一类集成电路。最早的“可编程逻辑器件”(PLD)是“可编程逻辑阵列”(PLA)。和门阵列一样,PLA 也是由“功能待定逻辑”(uncommited logic)组成的阵列。PLA 和“掩模编程”门阵列的区别是“用大电压(极性通常为负)可以对 PLA 进行配置”。图 13.2 是 PLA 的通用结构。PLA 有若干个输入(A,B,C)和若干个输出(X,Y,Z),一个“与”平面和一个“或”平面。图中显示了输入和乘积项 (P, Q, R, S)之间的连接以及乘积项和输出之间的连接。剩余的连接在编程时就被去除了。有些 PLA 是电可编程的,也可以使用紫外线照射进行编程。“可编程阵列逻辑”(PAL)扩展了 PLA,PAL 包含 12 个触发器。近年来,可编程器件变得更复杂了——包括“复杂可编程逻辑器件”(CPLD)和“现场可编程门阵列”(FPGA)。

设计流程

数字系统大多是“时序的”(sequential),即系统存在各种“状态”(state),输出取决于“现态”(present state)。早期的某些计算机设计是“异步的”(asynchronous),即输入稳定下来后就向新状态转移。数字系统的“同步”(synchronous)设计趋势已经有多年了。在同步系统中,状态的改变是由一个或多个时钟信号触发的。为了设计可靠的系统,人们定义了规范的设计方法。采用“分立门电路”设计(同步时序)数字系统,可按照如下步骤进行。

1. 写出设计规格。

2. 如果需要的话,将设计分解为若干个小部分,并为每一部分写出设计规格。

3. 根据设计规格,画出状态机转换图。该图包含系统中的每个状态、引起状态变化的输入条件和每个状态的输出。

4. 将状态数最小化。这是一个可选步骤,并不是在所有情况下都有用。

5. 为每个状态分配布尔变量。

6. 推导出次态逻辑和输出逻辑。

7. 对次态逻辑和输出逻辑进行优化,从而将所需的门数量减至最小。

8. 为逻辑门选择合理的布局,即集成电路中包含哪些逻辑门以及各个集成电路在印制电路板上的位置。

9. 设计集成电路之间的走线。

一般而言,步骤 1 和步骤 2 是不能忽略的。因为这里需要设计者创造性的工作。大多数数字设计书籍集中讨论步骤 3～步骤 7。步骤 8 和步骤 9 可以手工完成,而“元件布局”(placement)和“布线”(routing)是最早成功实现自动化的设计任务之一。如果将设计转换成计算机可读的形式,那么就可以在设计的不同阶段进行仿真。为了进行布局和布线,在步骤 7 中往往使用原理图输入程序,这样就可以输入电路的门级结构。原理图可以转换成为适合逻辑仿真器的形式。在步骤 9 完成之后,电路结构(包括互连线的电阻和电容引起的延迟)可以提取出来,然后进行再次仿真。

因此,在 ASIC 和 FPGA 上实现数字设计需要配置逻辑模块之间的连接。正像刚才提到的,我们不能够忽略步骤 1 和步骤 2,而步骤 8 和步骤 9 可以自动完成。使用 HDL 意味着可以将设计输入到 CAD 系统中,并可以在步骤 3 或步骤 4 进行仿真,而不是在步骤 7。所谓的“寄存器传输级”(RTL)综合工具可以自动完成步骤 6 和步骤 7。步骤 4 仍须手工完成。步骤 5 可以自动完成,并可以对某种状态分配结果进行快速评估。“行为综合”(behavioral synthesis)工具开始出现在自动化设计流程中的步骤 2 附近。图 13.3显示了基于 RTL 综合设计的整个设计流程。

第 14 课　现场可编程门阵列(FPGA)

FPGA 是什么?

现场可编程门阵列(FPGA)是一类数字“集成电路”(IC)。FPGA 由“可编程逻辑块”和块间的“可编程互联”组成。利用 FPGA,设计工程师可以编程实现各种各样的功能。

依据使用方式的不同,FPGA 可分为两类。一类 FPGA 只能编程一次,而另一类 FPGA 可多次编程。显然,只能编程一次的器件被称为“一次编程型”(OTP)FPGA。

FPGA 名称中的“现场可编程”是指(研发人员)“在现场”对 FPGA 编程(以实现所需功能),而不是由芯片制造商以“硬连线”方式实现器件的内部功能。也就是说,可以在实验室对 FPGA 进行配置,也可以对“已投入使用的”电子系统中的 FPGA 进行配置。如果系统中的某种器件具备可编程能力的话,那么就可称作“在系统可编程”(ISP)器件。

FPGA 引人关注的原因

数字集成电路的种类有许多,如“芝麻逻辑”(仅含少量简单、固定逻辑功能的小规模器件)、存储器件和“微处理器”(μP)。这里,我们特别关注的是“可编程逻辑器件”(PLD)、“专用集成电路”(ASIC)、“专用标准部件”(ASSP)和 FPGA。

为了便于讨论,这里的“可编程逻辑器件”(PLD)一词涵盖了“简单可编程逻辑器件”(SPLD)和“复杂可编程逻辑器件”(CPLD)。

本书第 2、3 章将全面、详细讨论 PLD、ASIC 和 ASSP。在此,我们只需了解一点——芯片制造商预先限定了 PLD 的内部结构。这种内部结构允许工程师“现场”编程实现各种功能。不过,和 FPGA 相比,PLD 包含的逻辑门比较少,而且 PLD 实现的功能也少得多、简单得多。

在该领域的另一端是 ASIC 和 ASSP,其逻辑门数量多达几千万,可实现的功能也极其庞大和复杂。ASIC 和 ASSP 的设计流程和制造工艺是相同的,都是为特定应用而定制的芯片。二者的唯一区别是 ASIC 是为某一家公司设计制造的,而 ASSP 面向的是众多的客户(此后当提及 ASIC 时,除非特别说明或与上下文不符,该术语同时也指 ASSP)。

尽管 ASIC 在大小(即晶体管数量)、复杂度和性能上都是最好的,但设计和制造一片 ASIC 却是极其耗时和昂贵的。ASIC 还有一个缺点,就是最终设计“固化”在芯片中、无法修改了。

而 FPGA 的位置处在 PLD 和 ASIC 的中间——FPGA 的功能可以“现场”定制,这一点像 PLD;FPGA 中几百万个逻辑门可以实现“以前只有 ASIC 才能实现”的庞大、复杂

功能。

开发基于 FPGA 的设计成本要比 ASIC 低得多(尽管 ASIC 在量产后非常便宜)。此外,基于 FPGA 的设计修改起来要容易,而且上市时间更快。FPGA 为众多具备创新能力的小型设计公司提供了生存机会。因为 FPGA 不仅用于“环境优越”的大公司进行系统开发,也便于“条件简陋”的小公司进行产品设计——即无须投入巨额的“一次性工程”(NRE)成本和购买昂贵的 ASIC 设计工具,几个工程师就能在 FPGA 测试平台上实现他们的软、硬件设想。在 2003 年,ASIC 和 ASSP 设计数量估计只有 1500～4000 个和 5000 个(这两个数字明显地逐年减少),而于同年启动的 FPGA 设计约有 45000 个。

FPGA 的用途

在 20 世纪 80 年代中期,当 FPGA 刚刚出现时,它主要用于实现“胶连逻辑”,“中等复杂度”状态机和比较有限的数据处理任务。在 90 年代早期,随着 FPGA 容量和复杂度的提升,当时的市场主要是通信和网络领域,FPGA 用来处理、传输大量数据。到 90 年代末,在消费产品、汽车和工业等领域,FPGA 的使用量出现了高速增长。

FPGA 常用于 ASIC 原型设计,或者作为新算法硬件实现的验证性平台。然而,由于开发成本低和上市时间短,FPGA 正越来越多地出现在最终产品中。实际上,几家大生产商已经拥有可以和 ASIC 竞争的 FPGA 器件。

在 21 世纪的头几年,几百万门的高性能 FPGA 就已经出现。它们用在嵌入式微处理器内核、高速输入/输出接口等应用中。现实情况说明:今天的 FPGA 几乎无所不能——它可以用来实现通信设备、软件无线电;它可以用来实现雷达、图像及其他数字信号处理应用;它可以用来实现构成“片上系统”(SoC)的全部软、硬组件。

具体地说,FPGA 正在占领以下四个主要市场:ASIC 和定制芯片、“数字信号处理”(DSP)、嵌入式微控制器和物理层通信芯片。此外,FPGA 为自己开辟了一个新市场——“可重配计算”(RC)。

1. ASIC 和定制芯片:如前面所讨论的,当今的 FPGA 用于过去是用 ASIC 和定制芯片才能实现的各种设计中。

2. 数字信号处理:高速数字信号处理早已由特制微处理器——“数字信号处理器”来实现了。然而,今天的 FPGA 包含了嵌入式乘法器、专用算术“布线”和大量片上 RAM——这些硬件为数字信号处理运算提供了方便。当上述特点结合 FPGA 强大的并行处理功能后,FPGA 的性能要超出最快数字信号处理芯片 500 倍以上。

3. 嵌入式微控制器:小型的控制功能通常由“微控制器”这种专用嵌入式处理器来实现。微控制器是一种低成本器件,在其处理器核周围集成了片上程序和指令存储器、计时器和输入/输出接口。而 FPGA 的价格不断下降,最小的 FPGA 器件也足以实现一个微处理器内核及一组定制的输入/输出功能。结果呢,对嵌入式控制应用而言,使用 FPGA 的吸引力与日俱增。

4. 物理层通信：长久以来，FPGA 用于实现通信芯片和高级网络协议之间的“胶连”逻辑。实际上，今天的高端 FPGA 可以容纳多个高速无线收发器——即可将通信和网络功能固化于单个器件中。

5. 可重配计算：利用 FPGA 的固有并行性和“可重配”特性对软件算法进行“硬件加速”。目前，许多公司正在建造基于 FPGA 的可重配计算引擎。这些引擎用于完成硬件仿真、密码分析、新药品研发等领域的计算任务。

第 15 课　VHDL 语言

VHDL 语言是什么？

VHDL 是一种为描述数字系统而设计、优化的编程语言。VHDL 具备多项适合描述电子元器件(从简单逻辑门到微处理器、定制芯片)行为的功能。这些功能可以精确描述诸如信号的上升时间和下降时间、门电路传输延迟、函数操作等电路行为的电气特性。利用既有的 VHDL 仿真模型作为构成模块，可以实现对更大型电路(采用原理图、方框图或系统级 VHDL 等描述方法)的仿真。

VHDL 是一种通用编程语言。和其他高级编程语言一样，VHDL 将复杂的设计概念表达为计算机程序；为了实现电路自动综合或系统仿真，VHDL 可将复杂电子电路的行为输入到一个设计系统中。和 Pascal、C、C＋＋一样，VHDL 也具有结构化设计的特点，并提供了丰富的控制指令集和数据表示法。和上述编程语言不同的是：VHDL 能够描述“并发”(concurrent)事件。这一点是非常重要的，因为(使用 VHDL 描述的)硬件在运行时就是“并发”的。

VHDL 最重要的应用之一是以所谓“测试平台”(test bench)的形式记录电路的性能指标。“测试平台”是验证电路时域行为的激励源及相应期望输出的 VHDL 描述。在 VHDL 工程项目中，“测试平台”是必不可少的，应该和其他电路描述一同创建。

一种标准语言

熟悉并掌握 VHDL 最具说服力的原因之一是：在电子设计领域，VHDL 是作为一种标准被采纳的。使用 VHDL 这样的标准语言，实际上保证了设计者将不必因新一代设计工具不支持已选择的输入方法，而放弃原有的设计思路，重新进行设计。使用标准语言意味着设计者可以更加充分地利用最新设计工具，可以了解数以千计其他工程师所积累的经验、知识——他们中的许多人也正在解决类似的问题。

VHDL 语言简史

VHDL 是指“超高速集成电路硬件描述语言”，它是在 20 世纪 80 年代早期作为美国国防部资助的一个高速集成电路研究计划的副产品而开发出来的。在这个计划中，研究人员反复遇到描述“庞大规模”(相对于当时技术水平而言)电路的难题和如何管理多个工程师小组参与进行的大型电路设计的难题。由于当时只有门级设计工具，人们很快就认识到需要性能更好、更加结构化的设计方法和工具。

为了应对这个挑战，由来自“国际商用机器公司”(IBM)、“德州仪器公司”(TI)和 Intermatrics 三家公司的工程师组成的一个小组和美国国防部签署了一份协议——要设计和实现一种崭新的、基于语言设计的描述方法。

VHDL 的最早版本(7.2)是在 1985 年公布的。1986 年，“电气与电子工程师协会”(IEEE)提出了语言标准化建议。经过由商业、政府和学术代表组成的小组对该建议进行了实质增强和改进之后，IEEE 于 1987 年完成了标准化工作。

该标准(IEEE1076－1987)就是今天市场上所有仿真和综合产品的基础。在 1994 年，IEEE 发布了 VHDL 的增强和更新版本(IEEE1076－1993)，VHDL 工具提供商马上做出反应——在其产品中加入了 VHDL 的最新特点。

虽然 IEEE 完整定义了 VHDL 语言，但是 VHDL 的某些方面使之很难实现完全可移植的设计描述(所谓“完全可移植的设计描述”是指该设计描述在使用不同工具生产商提供的仿真工具进行仿真时，会得到完全相同的仿真结果)。这个问题源于 VHDL 支持多种抽象数据类型，却没有很好地描述各种信号强度和一些常用仿真条件(如“未知”状态和“高阻”状态)。

在 IEEE1076－1987 被采纳后不久，为了使用户能够对复杂电路进行精确仿真，仿真器提供商开始提供非标准数据类型以对其进行改进。这就带来了问题，因为在某个仿真器中输入的设计描述在其他仿真环境下是不兼容的。VHDL 很快就变成非标准语言了。

为了避开非标准数据类型这个问题，IEEE 又研究出了另外一个标准。该标准(标准号为 1164)定义了一个包含标准 9 值数据类型定义在内的标准包(标准包是 VHDL 的特点之一，即允许将共用数据类型声明集中在一个外部库函数中)。标准数据类型称为 std－logic，而 IEEE 1164 标准包常被称作“标准逻辑包”。

IEEE1076－1987 标准和 IEEE1164 标准共同构成了当前广泛使用的 VHDL 标准。(IEEE1076－1993 正逐步进入 VHDL 主流，但它并没有为综合用户增添任何明显的新功能)。

1076.3 标准(常被称作“数值标准”或“综合标准”)定义了标准包和对与实际硬件相关的 VHDL 数据类型的解释。该标准于 1995 年底发布，目的在于替代由许多综合工具商创建的、在其产品中提供的定制(非标准)包。

IEEE1076.3 标准为综合用户做的和 IEEE1169 标准为仿真用户做的完全相同：在确

保兼容不同生产商工具的同时，提高 1076 标准的功能。1076.3 标准包含以下几项：

- IEEE1076 标准定义的比特型和布尔型数值的硬件解释文档，IEEE1164 标准定义的 std－ulogic 类型的硬件解释文档。
- 提供对 std－ulogic 类型数值进行“无关”(don’t care)或者“不确定”(wild card)测试的功能。该功能对综合尤为有用，因为用“无关”值来表达逻辑往往很有用。
- 对标准有符号和无符号算术数据类型的定义及其算术、移位、类型转换运算的定义。

“对仿真模型进行时序信息标注”是精确数字仿真的重要方面之一。VHDL1076 标准描述了一组可用于时序标注的语言特点。然而，它并未描述出一种在时序模型外表达时序数据的标准方法。

出于多种原因，将仿真模型的行为描述和时序技术规格分开是很重要的。Verilog HDL(VHDL 的竞争对手)的一项主要优势是：Verilog 包含了一种旨在于进行时序标注的功能。该功能(即标准延迟格式 SDF)允许以表格形式表示时序数据，并可以在仿真时将其包含进 Verilog 时序模型中。

在 1995 年的晚些时候，IEEE 公布的 IEEE1076.4 标准将这种功能以标准包的形式加入到 VHDL 中。推动这种标准化的主要原因是为了使 ASIC 等生产商更容易生成可同时应用于 VHDL 和 Verilog HDL 中的时序模型。因此，IEEE1076.4 和 Verilog SDF 底层数据格式是非常相似的。

何时采用 VHDL 语言？

为什么要用 VHDL 进行设计？原因可能有许多。当向 VHDL 工具提供商请教这个问题时，你首先得到的答案是“VHDL 将提高工作效率”。不过，这是什么意思呢？你真能指望使用 VHDL 比使用现有设计方法能更快完成手中的项目吗？回答是肯定的。不过，刚刚使用的时候很可能不会这样，只有以结构化方法应用 VHDL 才会这样。

VHDL 和结构化软件设计方法是相同的。当你使用结构化的、自顶向下的设计方法的时候，VHDL 才是有益的。当你超越了 VHDL 学习曲线，并积累了一定数量的可重用 VHDL 组件之后，工作效率很快就会出现真正的提高。

当你为增进项目小组成员间的交流而开始使用 VHDL 时，当你利用更强大的工具进行仿真和设计验证时，工作效率就会提高。此外，VHDL 允许在更抽象的层次上设计。你不再集中精力于门级实现，你可以去思考设计的行为功能。

VHDL 是如何提高工作效率的？是通过简化公用 VHDL 模块库的创建和使用。VHDL 使设计重用变得顺理成章。当发现了可重用代码的好处之后，你很快会发觉自己正在考虑采用 VHDL，从而使自己编写的语句更加通用。编写可移植代码将会成为自觉的行为。

使用 VHDL 的另外一个重要原因是 EDA 工具和目标器件技术的快速发展。使用

VHDL 这样的标准语言可以极大提高转用更高级工具(例如,从基本的低价位仿真器到更高级的仿真器)的可能性,而不必重新输入电路描述。使用标准输入方法也将提高将电路重定位于新型目标器件(如 ASICs,FPGAs 和 CPLDs)的能力。

第 16 课　数字信号处理的基本概念

我们说的话不是数字信号。数字信号是一种由“1”和“0”组成的、能用数学方法处理的语言。我们讲出的话是现实世界中的模拟信号。我们天天遇到的现实世界信号都是模拟信号——如声音、光、温度和压力。数字信号是模拟信号的数值表示。在数字世界里,对这些信号进行处理可能会更容易、成本更低。在现实世界中,我们可以通过“模数转换”将信号转换为数字信号,然后对信号进行处理;如果需要的话,用“数模转换器”将信号转换回到模拟世界中去。

模数转换和数模转换精要

对模拟信号进行采样是“模数转换”的关键第一步。这一步是由“采样保持电路”完成的,“采样保持电路”按照固定的“采样间隔”进行采样。

“采样间隔”的长度和“采样周期”的长度相同,而“采样周期”的倒数就是“采样频率”f_s。根据奈奎斯特采样定理,为了确保准确记录信号,最高频率为 W Hz 的信号(称为“带限信号”)每秒内必须采集至少 2W 个样本。如果不能满足这个最低要求,就会出现“混叠”失真。“混叠”会使高频信号出现在较低频段。为了保证不会出现“混叠”,在采样前要进行低通滤波。这个低通滤波器(称作“抗混叠滤波器”)将频率超过所选采样频率一半的信号全部滤除。

经过一段短暂的采集时间之后(其间采集到一个样本),采样保持电路将这个样本一直保持到下一个采样间隔的开始。A/D 转换器需要利用这个保持时间来生成模拟样本对应的码字。

A/D 转换器为每个模拟样值选择一个量化电平。一个 N 位转换器可以在 2^N 个可选量化电平当中进行选择。量化电平数越多,量化误差(量化电平和实际电平之差)越小。最大量化误差不会超过“量阶”Q 的一半。“量阶”的计算公式为 $Q = R/2^N$,其中 R 是模拟信号的满量程范围,N 是转换器所用的位数。相对于量化误差的信号强度使用“动态范围”和“信噪比”来度量。

数字信号是用一组竖线表示的,竖线顶部的圆圈标出样本选用的量化电平。A/D 转换器的比特率为 $N f_s$,这里的 f_s是采样率。

最后,要为每个数字化样本分配一个码字,这样就完成了 A/D 转换过程。A/D 转换

的结果就是得到一串数字比特流。数字信号处理就是对这些码字集合进行处理。

总结一下，A/D 转换过程由抗混叠滤波、采样、量化和数字化这四步组成。

数字信号处理完成之后，就必须进行 D/A 转换。首先，将码字转换为和码字所代表数字的大小成正比的模拟电压。这个电压值要在零阶保持器中保持到下一个码字出现，即需要保持一个采样间隔。这样，信号就是阶梯状的，信号中包含着频率超过 W Hz 的成分。D/A 转换的最后一步就是：使用平滑滤波器滤除那些频率超过 W Hz 的成分。

对于采样信号中的任意频率 f，其因采样而产生的镜像频率将出现在无数个频率处($k f_s \pm f$ Hz)。当采样率低于要求的奈奎斯特率(即 $f_s < 2W$)时，高频信号的镜像就会因为混叠而错误地出现在基带内(或奈奎斯特范围内)。尽管在通常情况下要避免这种“欠采样”，但是“欠采样”也有用途。例如，高频窄带信号的采样率可以是信号带宽的两倍，而不必是信号最高频率的两倍。信号的所有重要特征都可以从因采样而出现在基带的频谱副本中得到。由于信号频率和采样率之间的关系的不同，可能会出现“频谱反转”现象——基带频谱的形状和信号真实频谱的形状正好相反。

数字信号处理的实现技术

假如存在一种可以实现所有设计的“万能”处理器的话，电子行业的竞争就不会如此激烈了。对于多数电子设计而言，可用于实现所需功能的处理器技术不止一种。需要面临的问题是如何在预算范围内选择一种处理器技术，它不仅具备最佳的性能、大小、功耗特点，而且拥有易于快速开发的软件工具。经过了 20 年的发展之后，数字信号处理器仍然是具有竞争力的处理器。总之，数字信号处理器是信号处理的核心。

“数字信号处理器”(DSP)是一种速度极快、功能强大的微处理器。DSP 与众不同之处在于它能够实时处理数据。实时处理能力使 DSP 非常适合于那些不容许任何延迟的应用。例如，你用过两人无法同时讲话的手机吗？你必须等对方说完了才能开口。如果你们两人同时讲话，信号就会掉线——你听不到对方了。今天的数字手机使用了 DSP，这样人们就可以正常通话了。手机中 DSP 处理声音的速度非常快——你在说的同时就可以听到。和其他处理器相比较，使用 DSP 进行设计有如下几点优势：

- 可实现单周期“乘法－累加”操作
- 能够对实时性能进行模拟和仿真
- 灵活性好
- 可靠性高
- 有利于提高系统性能
- 有利于降低系统成本

不过，实现数字信号处理还可以选择其他技术。和 DSPs 相比，这些技术如何呢？

现场可编程门阵列

“现场可编程门阵列”(FPGA)具有“在系统可重新配置”的能力——对于开发需要多

次试用的应用，这种能力是巨大的优势，能提供快速的上市时间。由于生成了专用逻辑电路，所以对于指定任务 FPGA 能提供更好的性能。可是，FPGA 相当昂贵，功耗也往往要大大高于功能相近的 DSPs。在无线通信基础设备等设计中选用了 FPGA 技术，但通常是将 FPGA 和 DSPs 联合使用——这样做灵活性更大、价性比更好、系统功耗更低。

专用集成电路

通过设计，“专用集成电路”(ASIC)可以极好地完成某些特定功能，而且功率效率也可以设计得很好。不过，ASIC 无法进行现场编程；即使是在产品开发阶段，也无法对其功能进行修改和升级。因此，每次推出一个新产品都需要重新进行设计，并经历所有制造流程——这样做不仅造价昂贵，也不利于迅速上市。可编程 DSPs 与 ASIC 不同，可以不需要修改硬件而进行升级，升级只需修改软件程序——这样大大降低了系统成本，并可通过下载代码而实现(售后)功能增强。因此，在实时信号处理应用中，可编程 DSP 是系统的基础，而 ASIC 常用作系统的总线接口、“胶连逻辑”(glue logic)和(或)功能加速器。

通用微处理器

ASIC 是为特定用途优化的。与 ASIC 相反，“通用微处理器”(GPP)最适合于完成多种不同的任务。但在最终产品必须实时响应的应用中，或者必须在电池驱动下实现实时响应的应用中，由于实时性能较差、功耗大，所以 GPP 就被排除在外了。GPP 正在越来越多地被视为业界的庞然大物——为了适应不断变化的实时应用市场，不断增加的 PC 兼容性和台式电脑性能使 GPP 受到了拖累。因为人们越来越喜欢小巧的手持无线产品(其处理器功耗为毫瓦级，而不是瓦级)，所以 DSPs 就成为可编程技术的最佳选择。随着数字互联网设备变得更小、更快、更便于携带，这种趋势一定会保持下去。

第 17 课　数字信号处理器

只要是处理器，就可以完成数字信号处理任务。不过，专用“数字信号处理器”(DSPs)完成数字信号处理任务的效率和速度都是最好的。传统处理器遵循“冯·诺伊曼结构”，这种结构采用一个共享存储器，同时存储程序指令和数据；而 DSPs 采用的是“哈佛结构”或“改进哈佛结构”，这种结构包含多个程序和数据存储器以及访问这些存储器的多套总线。这样的设计就意味着从存储器取指令或取数据所花费的等待时间要少很多。实际上，至少可以同时取得一条指令和一个数据。这种任务的重叠称作“流水线”。除了多存储器和多总线之外，DSPs 都有高速乘法器、累加器、移位器，许多 DSPs 都有硬件支持“循环缓冲区”。“地址产生器”可以加速对寄存器寻址的存储器访问。

DSPs 分为定点和浮点两大类。定点 DSPs 使用固定的比特数来代表实数。二进制小数点的位置可以由编程人员决定，这个位置决定了可以表示的实数范围。可用精度会

随着表示范围的增加而下降，因为二进制小数点右边的比特数减少了。在 16 位数据中，可能出现的格式有 16.0，15.1，14.2，13.3，12.4，11.5，10.6，9.7，8.8，7.9，6.10，5.11，4.12，3.13，2.14 和 1.15。16 位定点数据格式的动态范围都是一样的，即 $20\log 2^{16} = 96.3$ dB。动态范围的计算方法为 20log(满量程范围/最小可分辨差别)。

浮点 DSPs 使用“尾数”(mantissia)和“指数”(exponent)来表示实数，这种方法很像科学记数法——将尾数和指数组合成一个 32 位数。浮点器件的动态范围是用 2^E 的最大值和最小值进行计算的，此处 E 是指数。对于 24 位尾数和 8 位有符号指数的表示方法而言，动态范围是 $20\log(2^{127}/2^{-128}) = 1535.3$ dB。大的动态范围意味着系统具备更大的表示很宽范围输入信号的能力，从很小的信号到很大的信号。

汇编语言是 DSPs 使用的命令语言。为了使常见数字信号处理任务的编程更加方便、高效，DSPs 往往采用专门指令。例如，多数 DSPs 都提供多功能指令，这些指令利用了 DSPs 的并行结构。DSPs 往往还提供高效的循环机制，因为许多数字信号处理运算都包含着大量的重复性操作。

为特定应用选择合适的 DSPs 不太容易。首先，需要明确选择定点器件还是浮点器件。一般而言，定点器件比较便宜、速度也快，而浮点器件更便于编程、更适于运算密集型算法。第二，DSPs 的数据宽度决定了它所代表数据的精度。速度也是要考虑的问题之一；速度不仅是指一秒钟包含多少个机器周期，还包括每个周期能够执行多少条指令以及这些指令中的每一条能够完成多少工作。一种评估 DSPs 最低要求的办法是：估计对每个到来的样本必须执行多少条指令。这个数乘以采样频率就得到了所需每秒指令数的最小值。

一种特定 DSP 提供的某些特定的软、硬件功能会使一种选择优于另一种选择，可用片内存储器的容量也是这样。有的时候，选择 DSPs 的理由是支撑硬件的匹配性很好，尤其是片上集成的 A/D 和 D/A 转换器。对于低级编程语言和高级编程语言，软件开发工具是否高效便捷也是经常要考虑的主要因素；是否能够得到第三方软件也是主要因素之一。而“成本”当然永远都是因素之一。事实上，被选中的 DSP 往往速度快、功能多、也符合资金预算要求。

购买 DSPs 有三种形式：购买内核、购买处理器和购买板级产品。在 DSPs 中，“内核”这个词是指处理器中运行关键任务的部分；它包括数据寄存器、乘法器、算术逻辑单元、地址产生器和程序定序器。一个完整的处理器需要将内核、存储器和外部接口组合起来。尽管内核设计和片内外设设计是分别进行的，但是却被制作在同一片硅片上；这样，处理器就能成为单片集成电路。

假定你要制造蜂窝电话并希望在设计中使用 DSPs，你很可能会以处理器的形式购买 DSPs；也就是说，你会购买一片包含内核、存储器和其他内部功能的集成电路。为了在产品中使用这片集成电路，你必须设计一个印制电路板，而 DSPs 将被焊接到这块电路板上，和其他电子元器件相连。这是使用 DSPs 最常见的方式。

假设你服务的公司要制造自己的集成电路。在这种情况下，你可能不会购买整个处理器，而仅会购买处理器的内核设计。在完成了相应的授权协议后，你就可以开始制造为特殊用途而定制的芯片了。这样做给了你选择片内存储器数量、数据收发方式和封装形式等方面的自由。在 DSPs 市场中，这种定制器件成为越来越重要的部分。

有几十家公司提供安装了 DSPs 的印制电路板。在这些电路板上，有附加存储器、A/D 和 D/A 转换器、EPROM 插座和多个处理器等。虽然有些电路板可用作独立工作的计算机，但大部分电路板配置成了主机（如 PC）插板的形式。制造这种电路板的公司被称为“第三方开发商”。寻找第三方开发商最好方式是询问你要使用的 DSPs 的生产商。查看 DSPs 生产商的网站；如果在那里找不到的话，就给生产商发个电子邮件。生产商非常乐于告诉你：谁在使用他们的产品以及如何与他们联系。

别忘了：DSPs 和其他微处理器的界限并不是很清晰。例如，让我们看一看 Intel 是如何描述其奔腾处理器新增的 MMX 技术的：“为了高效操作和处理视频、音频和图形数据，Intel 工程师新增了 57 条功能强大的指令。这些指令面向的是多媒体操作中经常出现的、高度并行和重复的程序。”

在将来，我们一定会看到更多的数字信号处理功能融合到传统微处理器和微控器当中。推动这种变革的强大动力之一就是网络和其他多媒体应用。此类应用的发展速度非常快，以至于 20 年后的 DSPs 很可能也会变成“传统”微处理器了。

第 18 课　数字信号处理和模拟信号处理

信号既可以使用模拟技术进行处理（模拟信号处理 ASP），也可以使用数字技术（数字信号处理 DSP）进行处理，还可以使用模拟、数字混合技术进行处理（混合信号处理 MSP）。在有些情况下，选择哪一种技术非常容易明确；而在另一些情况下，却无法一下子明确选择哪种技术，需要多次衡量后才能做出最终决定。

说到 DSP，它和传统计算机数据分析的区别在于 DSP 在实时进行滤波、FFT 分析和数据压缩等复杂信号处理时的速度和效率。

“混合信号处理”这个词包含“模拟处理和数字处理都是系统一部分”这样的意思。系统的实现形式可能是一块印制电路板，也可能是一片集成电路。在这个宽泛的定义之下，ADC 和 DAC 也可视作“混合信号处理器”，因为它们当中既实现了模拟功能，也实现了数字功能。“超大规模集成”（VLSI）技术的新近进展允许在一片芯片内既可实现模拟处理，也可实现复杂的数字处理。DSP 本身的特性就表明：它可以实时完成这些功能。

模拟信号处理和数字信号处理

今天，工程师们面临着这样的挑战：为了解决手头的信号处理任务，怎样选择模拟技术和数字技术的恰当组合。单靠纯数字技术是不能完成处理现实世界模拟信号的任务的，因为所有传感器(话筒、热偶、压力感应器、压电晶体、磁盘驱动器头等)都是模拟的。因此，为了使传感器的输出信号能够进一步地进行信号处理(不论模拟信号处理，还是数字信号处理)，就需要某种形式的"信号调理电路"。实际上，信号调理电路就是模拟信号处理器，它能够完成乘法(增益)、隔离(测量放大器和隔离放大器)、噪声环境中的信号检测(高共模抑制比的测量放大器)、动态范围压缩(对数放大器、对数数模转换器和增益可编程放大器)和滤波(包括无源滤波器和有源滤波器)等功能。图 18.1 展示了实现信号处理的几种方法。图的最上方是纯模拟方法。接下来是 DSP 的方法。注意：一旦决定使用 DSP 技术，接下来就要决定将 ADC 置于信号通道中的哪个位置。

一般而言，ADC 越靠近实际的传感器，ADC 将要担负的信号调理任务就越重。ADC 复杂性的提高可表现为如下几种形式：提高采样率、加宽动态范围、提高精度、抑制输入噪声、对输入信号进行滤波和"增益可编程放大器"(PGAs)、片内参考电压等。所有这些都增加了 ADC 的功能、简化了系统。有了当前高精度、高采样率的数据转换技术，ADC/DAC 在集成更多调理电路方面取得了重大进展。例如，在测量领域，人们能够得到具有内置 PGAs 的 24 位 ADC，它可以对满量程的 10 mV 桥式整流信号直接进行数字化，而不需要进一步调理。在音频和声频领域，出现了完整的编解码器(Codec，又称"模拟前端")，它拥有数量充足的片内模拟电路，从而最大程度地减少了对外部调理器件的需求。在视频领域，CCD 图像处理等应用也有了模拟前端。

实例分析

为了理解 DSP 的功能，这里给出一个实例。对一个模拟低通滤波器和一个数字低通滤波器进行比较，二者的截止频率都是 1kHz。这个数字滤波器是通过图 18.2 所示的典型采样数据系统来实现的。注意：图中有几个隐含的要求。首先，假定 ADC 和 DAC 有足够高的采样率、精度和足够大的动态范围，能够精确地进行信号处理。第二，DSP 必须足够快，能够在采样间隔 $1/f_s$ 内完成全部运算。第三，在 ADC 输入端和 DAC 输出端依然需要模拟滤波器来完成"抗混叠"和"抗镜像"的功能，但是对其性能的要求不像对数字滤波器那么严格。假定这些条件都能够得到满足，那么下面就对这两种滤波器进行一下比较。

两个滤波器要求的截止频率都是 1kHz。模拟滤波器用 6 阶切比雪夫 I 型滤波器(通带有波纹，阻带没有波纹)实现，其频率响应见图 18.3。在实际应用中，这个滤波器很可能是用 3 个二阶滤波器实现的，每个二阶滤波器需要一个运算放大器、一些电阻器和一些电容器。现代滤波器"计算机辅助设计"(CAD)软件能够比较容易地进行六阶滤波器设

计，但是“0.5 dB 波纹”的指标还是需要对元器件进行精心选择和匹配才能实现。

另一方面，图中的 129 抽头 FIR 滤波器仅有 0.002 dB 的通带波纹，而且具有线性相位特性，其滚降特性也陡峭得多。实际上，这个滤波器是无法用模拟技术实现的！数字滤波器的另一个优势是不需要元器件匹配，而且对于漂移不敏感(因为时钟频率是受晶体控制的)。这个 129 抽头滤波器需要进行 129 次“乘法－累加”运算，才能计算出一个输出样本。为了保证实时运行，处理过程必须在采样间隔 $1/f_s$ 内完成。在本例中，采样频率为 10 kSPS，因此可利用的处理时间有 100 μs(假定没有重大额外开销的话)。多数 DSPs 能够在一个指令周期内完成整个全部的“乘法－累加”处理(和其他滤波器所需的功能)。

因此，一个 129 抽头滤波器需要指令速率大于 129/100 μs = 1.3 MIPS。我们可以得到指令速率大大超过 1.3 MIPS 的 DSPs；因此，在这个应用中，DSPs 肯定不是限制性因素。

在实际应用中，在评价模拟滤波器和数字滤波器(或模拟信号处理和数字信号处理)的时候，肯定还要考虑许多其他因素。为了实现期望的功能和充分利用两种技术各自的优势，多数现代信号处理系统都采用了模拟技术和数字技术相结合的方式。

第 19 课　高保真音频

高保真音响爱好者对声音质量的要求极高，所有其他因素都被视为次要的。假如要用一个词来描述这种心理状态的话，那就是“过分”。高保真音响系统不是设计得刚好满足人类的听觉需求，而是超越了人类听觉的极限。这是唯一能够确保再现音乐无任何失真的方法。CD 带给世人数字音频的享受。这是音乐领域里的巨大变化，CD 系统的音质远远超过了传统的唱片、磁带。

图 19.1 显示了高倍显微镜下的 CD 盘面。盘面的主要部分是闪闪发光的(反射光线)，数字信息的存储形式是激光烧制在盘面上的一组暗坑。信息排列在由外到内的一条螺旋线形轨道上(这一点和唱片是一样的)。当从 CD 外侧向 CD 内侧读取信息时，CD 的转速从 210 转/分变到 480 转/分，从而保持扫描速率为恒定的 1.2 m/s(唱片的转速是固定的，如 33 转/分、45 转/分或 78 转/分)。播放时，光传感器检测表面是否反射光，并产生相应的数字信息。

从图 19.1 显示的几何尺寸可以看出，CD 每平方微米存储 1 位信息(相当于每平方毫米存储 1 百万位信息)，共存储 1.5 亿位信息。这个尺寸和集成电路制造中的特征尺寸大致相同。光是无法聚焦到小于大约 1/2 波长(或 0.3 μm)的，这是光的特性之一。由于集成电路和光盘都是使用光学方法制造的，因此在 0.3 μm 以下光的模糊性限制了可用特征尺寸的大小。

图 19.2 是一个典型 CD 播放系统的组成框图。原始数据率为每秒 430 万位，相当于每 0.28 μm 轨道长度存储 1 位。然而，这和指定的 CD 几何尺寸发生了矛盾。每个凹坑的长度不准小于 0.8 μm，也不准大于 3.5 μm。换言之，每个二进制“1”必须是 3 到 13 个“1”中的一部分。这样做的好处是降低了光学读取造成的错误率。但是，如何让二进制数值遵循这种奇怪的安排?

答案是使用一种称作“8－14 调制”(EFM)的编码方法。这种方法不是直接将一个字节的数据存储到光盘上，而是使用查找表将这个 8 位数据转换为一个 14 位的数据。这种 14 位数据具备所需的特性而被存储在光盘上。播放时，从光盘读取的 14 位二进制数据通过反向搜索 EFM 查找表找到对应的 8 位数据。

除了 EFM 之外，数据以二级“里德 · 索罗蒙编码”方式进行编码。这种编码方法对左、右声道和检错、纠错数据进行组合。播放时检测到的错误，可以使用编码机制中的冗余数据进行纠正，也可以用邻近样本插值的方法去除，还可以用“样本置零”的方法使之“静音”。这些编码机制使得数据率增至原来的三倍——即 1.4 Mbps 的立体声音频信号将以 4.3 Mbps 的数据率存储在光盘上。

在完成解码和纠错后，音频信号将用 44.1 kHz 采样率的 16 位样本来表示。在最简系统中，信号通过 16 位“数模转换器”(DAC)，之后由低通模拟滤波器进行滤波。然而，这样做就需要高性能的模拟器件去实现如下滤波功能：让 20 kHz 以下的频率成分通过，而阻止 22.05 kHz(采样率的一半)以上的频率成分通过。“多采样率”技术是更常用的方法——在 DAC 前将数据转换为更高的采样率。常用的倍增因子是 4，即将 44.1 kHz 转换至 176.4 kHz。这就是所谓的“插值”，这个过程可以理解为两个处理步骤(尽管实际上不可能用这种方式实现)。第一步，在原有数据的每两个样本之间，插入三个零值样本，从而产生了更高的采样率。在频域，这样做会产生将 0～22.05 kHz 频谱复制三次的作用；这三个频谱副本分别位于 22.05～44.1 kHz，44.1～66.15 kHz 和 66.15～88.2 kHz。第二步，用一个高效数字滤波器去掉这些新增的频率分量。

提高采样率就减小了采样间隔，DAC 生成的信号就更平滑了。这个信号中依然包含着 20 Hz～20 kHz 的频率成分，而奈奎斯特频率已经提高至原来的四倍。这意味着模拟滤波器仅需使 20 kHz 以下的频率成分通过，而阻止 88.2 kHz 以上的频率成分。通常，一个三阶贝塞尔滤波器就可完成这个功能。

因为样本数是原先的 4 倍，所以代表样本的位数可以由 16 位降至 14 位，但却不会降低声音质量。用于补偿 DAC 零阶保持器的 sin(x)/x 校正既可以是模拟滤波器的一部分，也可以是数字滤波器的一部分。

多于一个声道的音频系统称为“立体声系统”(stereo 源于表示“立体”、“三维”的希腊单词)。多个声道从不同方向向听众发送声音，这样能更准确地重现原来的音乐。单声道系统播放出来的音乐往往听上去生硬而又平淡。相比之下，一个好的立体声音响系统会使听者感到似乎演奏家们距离自己只有几步之遥。从 20 世纪 60 年代以来，高保真音乐

就已开始使用双声道(左声道和右声道),而电影也已开始使用四声道(左声道、右声道、中央声道、环绕声道)。在早期立体声录音中,往往只能在一个声道听到歌唱家的声音。很快,这就发展成为更复杂的“混音”(Mix-down)技术,录音棚中来自多个麦克风的声音被合成到两个声道中。Mix－down 是一种旨在为听者提供“现场感”的技术。

电影中四声道立体声称作“杜比立体声”,其用于家庭的版本称作“杜比环绕 ProLogic”(Dolby 和 ProLogic 是杜比实验室及经其认证的产品所使用的商标)。四声道声音被编码成标准的左、右声道,这样常规的双声道立体声系统也可播放音乐。杜比解码器用于在播放时重现四声道声音。在电影或电视荧屏两侧的扬声器中传出的两个声道的声音和常规双声道声音是很相似的(见图 19.3)。产生中央声道的扬声器常被安置在荧屏的正上方(或者正下方)。目的是不论观众或听众的座位如何,重现语音和其他视觉相关声音都固定在荧屏中央。环绕扬声器放置在听众的左右,而在大型礼堂中可能有多达 20 个环绕声道;环绕声道仅包含中间频率(如 100 Hz 到 7 kHz),而且有 15～30 ms 的延迟。这个延迟使听众认为声音是来自荧屏而不是来自两侧。也就是说,听众首先听到来自荧屏的声音,然后听到来自两侧经过延迟的声音。听众的大脑会把延迟信号当作是墙的反射,而将其忽略。

第 20 课　音频压缩

CD 音质的音频信号需要 1.411 Mbps 传输带宽。如果要使网络传输成为现实,显然需要进行实质性压缩。各种各样的音频压缩算法开发出来了,MPEG 音频算法或许是最流行的一种。MPEG 算法分为三层,其中 MP3(MPEG 音频第三层)最有效、最著名。在互联网上,人们可以获得大量 MP3 格式的音乐。不过,并非所有这些音乐都是合法取得的;因此,就出现了大量由艺术家和版权所有人提起的诉讼。

MP3 是 MPEG 视频压缩标准中的音频部分。音频压缩可以通过两种方式实现。在“波形编码”中,用“傅里叶变换”这种数学方法将波形信号变换为频率分量。每个分量的幅度采用最小方式进行编码,目的是用尽可能少的比特数在另一端准确重现波形。另外一种方法就是“知觉编码”。这种方法利用了人类听觉系统的缺陷,采用让人耳听不出来差别的方式对信号进行编码,尽管在示波器上观看重放波形差别很大。知觉编码技术是建立在“心理声学”基础之上的。“心理声学”研究的是人类感知声音的方式。而 MP3 建立在知觉编码基础之上。

知觉编码的一个关键特性是:一些声音可以掩蔽另一些声音。想象一下:在一个温暖的夏日,你正在收听长笛演奏会的实况广播。突然,附近有一组工人开动了手提钻,并开始切割街道路面。谁也听不见长笛的声音了,它的声音被手提钻的声音掩盖了。从传输

的角度看，现在只要对手提钻所在频率进行编码就足够了，因为听众再也无法听到长笛声。

这种现象称作“频率掩蔽”——某个频率上响度较大的声音能够掩盖另一个频率上响度较小的声音。假如响度大的声音不存在的话，这个响度小的声音本来是可以听得到的。事实上，即使在手提钻停止工作后的一小段时间内，还是听不到长笛的。因为手提钻开始工作的时候，人耳调低了其增益；而将人耳增益再次调高需要一段时间。这个效应称作“暂时掩蔽”。

为了使这些效应更加量化，想象一下实验 1。在安静的房间里，一个人将耳机连至计算机的声卡上，计算机产生一个小功率 100 Hz 纯净正弦波，正弦波的功率在缓慢增加。这个人被告知：当他听到这个音的时候，就敲击一下键。计算机记录着当前的功率电平，然后在 200 Hz，300 Hz 和其他频率上重复这个实验，直到人耳听觉的极限。在对多个实验结果进行平均之后，就得到了一张和图 20.1(a)相像的有关“具备多大功率的单音才能被听到”的对数－对数图。

从该曲线中，直接可以得出如下结论：对“功率在可听门限以下”的频率成分进行编码是绝对没有必要的。例如，在图 20.1(a)中，如果 100 Hz 频率信号的功率为 20 dB，那么这个信号就可以从输出信号中略去，却不会出现可觉察的音质降低；因为 100 Hz 上 20 dB 的功率在可听电平之下。

现在考察一下实验(2)。计算机再次运行实验(1)，但这一次用一个固定幅度正弦波(比如说 150 Hz)叠加到测试频率上。我们发现频率位于 150 Hz 附近的信号的可听门限提高了(如图 20.1(b)所示)。

这次观测可以得到如下结果：通过跟踪哪些信号会被邻近频带更强的信号屏蔽，我们就可以在编码信号中忽略更多的频率成分，从而节省了数据位数。在图 20.1 中，输出完全可以忽略 125 Hz 信号，而没有人能听出差别；甚至某个频带上的一个强信号消失了，但由于“暂时掩蔽”特性，在接下来的一段(人耳)恢复期内也可以忽略被屏蔽的频率。MP3 算法的核心就是利用傅里叶变换获得声音在每个频率上的功率，然后只输出那些不被屏蔽的频率，并用尽可能少的比特数对其进行编码。

有了这些背景知识，我们现在可以看一下 MP3 编码是如何进行的。音频压缩使用 32 kHz、44.1 kHz 或 48 kHz 对波形进行采样。采样可以是单声道的，也可以是双声道的，并且可以选用如下配置之一：

1. 单声道(一路输入信号流)
2. 双单声道(如一路英语音轨和一路日语音轨)
3. 非联合立体声(各声道分别压缩)
4. 联合立体声(充分利用声道间冗余)

首先，要选择输出数据率。MP3 能将一张摇滚乐 CD 压缩至 96kbps，而几乎没有可觉察的音质下降；即便是摇滚乐爱好者也听不出音质下降。对于钢琴音乐会而言，至少需

要 128 kbps。两个数据率的不同源于摇滚乐的“信噪比”远远高于钢琴音乐会（当然，这是在工程意义上讲的）。也可以选择更低的输出数据率，但在音质上会出现一些下降。

在这之后，样本以 1152（约 26 ms）为一组进行处理。每组样本首先通过 32 个数字滤波器，从而得到 32 个频带。同时，输入信号进入心理声学模型以决定被屏蔽的频率。下一步，32 频带中的每个频带进一步变换以得到更好的频谱分辨率。再下一步，将可用的比特数分配给每个频带，谱功率大的“未屏蔽”频带分配到较多的比特数，谱功率小的“未屏蔽”频带分配到较少的比特数，而完全被屏蔽的频带不分配比特数。最后，用“哈夫曼编码”方法对这些数据进行编码。“哈夫曼编码”将短码字分配给出现频繁的数据，而将长码字分配给出现不频繁的数据。

事实上，（相关内容）还不止这些。还有各种不同的技术用来进行噪声消除、抗混叠和声道间冗余的挖掘，但是这些内容已超出了本书的范围。

第 21 课　第三代移动电话：数字话音和数据

移动电话的未来是什么样的呢？让我们大致预测一下。有几个因素推动着这个行业的发展。首先，在固定网络中，数据业务量不但超过了话音业务量，而且还在以指数速度增长，话音业务量却基本保持不变。许多业内专家预计：在移动设备中，数据传输也将很快赶超话音业务。第二，电话、娱乐和计算机均已数字化，三者正在迅速融合。许多人盼望着通过轻巧的便携设备来完成电话、CD 播放器、DVD 播放器、电子邮件终端、Web 接口、游戏机、文字处理器等众多功能；而所有这些功能均可以在世界范围内以很高的带宽（数据率）无线接入互联网。这种设备及其连接方法正是第三代移动通信的研究内容。

1992 年，“国际电联”（ITU）更具体地描述了上述构想，并公布了称作“IMT－2000”的蓝图。其中，IMT 代表“国际移动电信”；而数字“2000”有以下三个含义：（1）是指开通这项业务的年份；（2）是指该业务的工作频段（以 MHz 为单位）；（3）是指该业务将拥有的带宽（以 kHz 为单位）。

以上三项指标，ITU 都没有做到。到了 2000 年，什么也没有实现。ITU 建议各国政府预留 2 GHz 附近的频谱，以便设备能够在国家间实现无缝漫游。后来，人们认识到：对于移动性太强的用户，2 Mbps 在目前是不现实的（因为很难实现快速的越区切换）。比较现实的做法是将 2 Mbps 用于室内用户（这将会和 ADSL 直接交锋），384 kbps 给步行用户，144 kbps 用于车载用户。

IMT－2000 网络提供给用户的基本业务有：

1. 高质量的话音传输

2．消息（取代电子邮件、传真、短信、聊天等）

3．多媒体（播放音乐，观看视频、电影和电视等）

4．互联网访问（网上冲浪，包括含音、视频的网页）

IMT－2000的附加业务可能有视频会议、遥现、多人游戏和移动商务。不仅如此，还可以在世界范围内瞬时地、高质量地提供所有这些业务。在没有地面网的地方，可以通过卫星实现自动连接。

ITU提出的IMT－2000是一种全球适用的技术。生产商可以生产世界任何地方均可销售、使用的设备（就像CD播放器和计算机一样，而不像现在的移动电话和电视）。统一的技术也将简化网络运营商的工作，并可以鼓励更多的人使用。像Betamax和VHS在录像机问世之初时的格式之争，对商业并没有什么好处。

业界提出的3G方案不止一个。经过筛选，最终得到两个方案。一个是爱立信（Ericsson）公司提出的“宽带CDMA”（W－CDMA）。W－CDMA采用的是“直接序列扩频技术”（DSSS），其工作带宽为5 MHz。虽然W－CDMA并不向下兼容GSM，但却可以和GSM网络一同工作。W－CDMA具备以下特性：当通话者从W－CDMA小区进入GSM小区时，通话不会中断。W－CDMA得到欧盟的大力推介，并称之为“通用移动电信系统”（UMTS）。

高通（Qualcomm）公司提出的CDMA2000是3G方案的另一个竞争者。CDMA2000采用的也是“直接序列扩频”技术。它基本上就是IS－95的扩展，并向下兼容IS－95。CDMA2000也使用5MHz带宽，但它并没有设计成和GSM一同工作，因此无法将通话切换到GSM小区。此外，CDMA2000和W－CDMA在码片速率、帧长、所用频谱、时间同步方式这些方面也存在差异。

假如爱立信和高通的工程师们被请到一起并告知完成某个共同设计，他们很可能会这样做。毕竟，这两个系统的基本原理都是5MHz信道上的CDMA，而且没有人愿意追求码片速率的不同。但问题的真正难点不是工程方面的。欧洲想要一个能和GSM一同工作的系统，而美国想要一个和美国广泛使用的系统（IS－95）相兼容的系统。双方也支持各自的本地公司（爱立信公司总部在瑞典，而高通公司在加州）。最终的结果是：爱立信公司和高通公司在各自CDMA专利权问题上卷入了一起又一起的官司。

1999年3月，两个公司的官司解决了——爱立信公司同意购买高通公司的基础设施。双方也达成了“同意建立单一3 G标准”的协议。不过，这个标准包含了多个不兼容可选项。在很大程度上，这么做的目的只是为了掩饰双方在技术上的分歧。尽管存在这些争执，3 G设备和业务在未来的几年内很可能就会出现。

在等待3 G之争停止的同时，有的运营商向3 G（有时被称作2.5 G，或许将其称为2.1 G更准确些）的方向上迈出了谨慎的一小步。“增强数据率GSM”（EDGE）就是这样的系统；EDGE在GSM的基础上提高了每波特包含的比特数。由此带来的问题是：每波特比特数越多意味着每波特误码率越高。因此，在调制和纠错方面，EDGE采取了与

GSM 不同的 9 项机制；这种不同体现为：将带宽的多少用于纠正高速传输引起的误码。

“通用分组无线业务”（GPRS）是另一个 2.5 G 方案。GPRS 是一种建立在 D-AMPS 或 GSM 上的分组网络。它允许移动站在运行话音系统的小区中发送、接收 IP 分组。当 GPRS 运行时，要将一些频率上的若干时隙预留用于分组传输。时隙的数量和位置可由基站进行动态管理，这取决于小区内话音业务量和数据业务量之比。

可用时隙被分成几个不同用途的逻辑信道。基站决定逻辑信道和时隙之间的映射关系。一个逻辑信道用于将分组从基站下载到某个移动站，而每个分组都指明了自己要去谁那里。为了发送一个 IP 分组，移动站向基站请求一个或几个时隙。如果这个请求正确抵达基站，基站将声明分配给用于发送该分组的频率和时隙。一旦分组抵达基站，它就通过有线连接被转交给互联网。因为 GPRS 仅仅是在已有话音系统之上的覆盖，所以它最多是在 3 G 到来前的一种权宜之计。

目前，一部分研究人员已经着手研究 4G 系统了。他们提出的 4 G 系统具备如下特征：大带宽，无处不在（任何地点均可接入）、和有线网络（尤其是 IP）的无缝集成、资源和频谱的自适应管理、软件无线电和高质量多媒体业务。

第 22 课　数字图像基础

数字图像分辨率

一幅数字图像是由多行多列“像素”（pixel）组成的。对灰度图像而言，每个像素被赋予一个代表其灰度级的数字。一幅图像拥有的像素越多，可用灰度级越多，这幅图像的分辨率就越高。下图是一幅 8 比特图像，可用灰度级为 $2^8=256$。行列数为 808×562。

直方图

如图 22.2 所示，一幅数字图像的灰度级可以用“直方图”（Histogram）来概括。直方图以柱状图的形式反映了在图像中出现的每个灰度级对应的像素数量。当一幅图像只用了可用灰度级的一小部分时，可以使用“直方图均衡”的方法将灰度级的使用扩展到整个可用范围。“直方图均衡”对灰度分布重新安排，提高了图像对比度。

图像加减

对于数字图像，可以对每个像素进行加法和减法。两幅图像相加可以把两组物体合到一幅图像之中。此外，对同一场景的几幅“含噪图像”进行平均可以降低噪声的影响。而图像相减可用于去掉图像中不需要的背景。将拍摄时间上前后相继的两幅照片相减，

将会显示出两幅照片中出现变化的地方。

当两幅图像进行加、减法的时候，结果矩阵往往包含不正确的灰度值。例如，在一幅8比特图像中，某个像素灰度级为127；在另一幅8比特图像中，相应像素的灰度级为201；那么，(图像相加后)“和像素”的灰度级是330。这个数值超出了8比特图像灰度级的有效范围——8比特图像能够包含的灰度值是0～255。这两幅图像相减的时候，“差像素”为−72，这个数值也超出了合法范围。鉴于上述情况，大多数算术运算之后都要进行数据缩放。可以按照下面的方法将数据缩放到范围[0，GSLmax](公式略)。

图像滤波

通过使用卷积算子进行二维卷积运算可以实现对数字图像的滤波。一幅N×N图像被M×M卷积算子滤波时，“边缘效应”会使图像的每一边都失去(M−1)/2行和列的像素。低通滤波器会使图像变得模糊，而高通滤波器突出了图像的灰度锐变，边缘滤波器对图像边缘进行定位。

膨胀和腐蚀

“膨胀”是向图像中所有物体添加一层像素，“腐蚀”是从所有物体上去掉一层像素。在膨胀之后进行腐蚀，就可以填补(由边缘检测识别产生的)破碎边界中的缝隙。相反，在腐蚀之后进行膨胀，就可以去除图像中的噪声点。要实现可靠的物体边界识别，首先就要检测图像边缘。利用边界信息，可以计算出周长、面积等形状特征，这些特征可用于“物体分类”。

扭曲和变态

“扭曲和变态”是应用于娱乐、医学成像领域中的数字图像技术。“扭曲”对图像中的物体进行拉伸或重塑，而“变态”将一幅图像变换成另外一幅图像。这种变换的实现过程如下：在源图像中标记控制点、控制线或控制三角形，并为其选择在目标图像中的新位置。然后，通过对控制元素的位置、颜色进行平滑变换，实现两幅图像间的转换。那些没有明确标记为控制元素的像素的位置、颜色由离其最近的控制元素的位置、颜色来确定。

图像频谱

“快速傅里叶变换”(FFT)可以用来分析数字图像的频谱。二维频谱包括“幅度谱”和“相位谱”，这和一维频谱是相同的。“相位谱”携带着图像中目标位置的信息。当幅度全部设置为1时，仅靠“相位谱”也能显示出原始图像的大概。而当相位全部设置为零时，单靠“幅度谱”却什么都显示不出来。

“图像频谱”是“计算机断层造影”(CT)、“核磁共振”(MRI)这两种扫描成像技术的基础。CT技术使用X射线从不同方向对目标的某个平面进行扫描。MRI扫描取决于物体在

变化磁场中表现出来的磁学特性。这两种扫描方式都允许对三维物体进行非侵入性探测。

图像压缩

部分由于互联网的原因，数字图像比以往更频繁地从一地传送到另一地。为了节省时间和带宽(空间)，(和其他文件一样)经常需要在传输前对图像文件进行压缩。“无损压缩方法”压缩了文件，却没有损失原始文件的任何信息——重建文件和原始文件是完全相同的。“有损压缩”是指原始文件中的有些信息永远丢失了；不过，重建文件和原始文件非常相近。“压缩比”是指原始文件大小和压缩文件大小之比。

“行程编码”(RLE，或称“游程编码”)是一种简单的压缩方法。“行程编码”方法将一个数字超过三次的重复次数附加在这个数字后面。“哈夫曼编码”(或称“霍夫曼编码”)是另外一种编码方法。这种方法用较短的代码来表示经常出现的信号，而使用较长的代码来代表不经常出现的信号。

JPEG 是一种极为常见的图像压缩方法。JPEG 使用“离散余弦变换”(DCT)将一幅图像 8×8 小块中的大部分信息集中到少数几个重要系数上。然后，JPEG 使用行程编码和哈夫曼编码对图像进一步压缩。

第 23 课　数码相机

和常规相机一样，数码相机也是通过一组透镜将光线聚集并形成景物的像。不过，数码相机并不将光线聚集在胶片上，而是将光线聚集在半导体器件上，这些半导体器件以电信号的形式存储光信号。然后，微处理器把电信息转换成数字数据。

图像传感器

多数数码相机采用“电荷耦合器件”(CCD)图像传感器。一些低端相机采用 CMOS 技术。尽管将来 CMOS 传感器一定会提高性能，并将更受欢迎。但在高端数码相机中，CMOS 传感器不大可能替代 CCD 传感器。

CCD 器件采用了一种特殊的制造工艺，能够无畸变地将电荷传送出去。正是因为这种工艺，CCD 器件的“保真度”和“光敏性”都非常好。与此相反，CMOS 器件采用的是完全标准的制造工艺，这种工艺和用来制造大多数微处理器的工艺是相同的。由于制造上的差异，CCD 传感器和 CMOS 传感器之间存在明显不同。

1. CCD 传感器可以产生高质量、低噪声的图像。传统的 CMOS 传感器更易受噪声影响。

2. 传统 CMOS 传感器功耗很低，而 CCD 传感器的功耗要比同等 CMOS 传感器多

100 倍。

3. CMOS 芯片可以在任何标准硅片生产线上制造。因此，和 CCD 传感器相比，CMOS 传感器极为便宜。不过，CCD 传感器进行大规模生产的时间很长了，所以其制造技术更加成熟。

分辨率

相机能够捕捉到的细节量称为“分辨率”，其衡量单位是“像素”。当前数码相机使用的典型分辨率如下：

1. 640×480 像素：这是多数低端相机的分辨率。如果你打算将图片以电子邮件形式发送给朋友或放在网站上，这个分辨率就可以了。其像素总量为 307,000。

2. 1216×912 像素：如果你打算打印图像，这个分辨率很好。其像素总量为 1,109,000。

3. 1160×1200 像素：这是高分辨率。按照这个分辨率拍到的图像在放大打印时（如 8×10 英寸）的效果也很好。其像素总量接近 2 百万。今天，可以找到像素总量达 1200 万的相机。

在喷墨打印机上，什么样的图片分辨率会有最佳效果呢？“柯达公司”（Kodak）推荐了不同打印尺寸下的最小分辨率。（表略）

色彩捕获

为了获得全彩图像，多数传感器使用滤光技术将光分解为三基色。一旦将这三种光全部记录下来，就可以将其相加而产生整个可见光谱。在数码相机中，有几种记录三基色光的方法。高质量相机使用三个独立传感器，每个传感器上方都有一个不同的滤光片。在相机中放置分光镜，就可将光线引向不同的传感器。我们可以把光线进入相机的情形想象成水流经管道。使用分光镜就像将等量的水分配给不同的管道。每个传感器看到的图像是完全一样的；不过，由于滤光片的存在，每个传感器仅响应三基色中的一种颜色。这种方法的优势是：在每个像素位置上，相机记录了每一种基色。但可惜的是，使用这种方法的相机又大又贵。

另外一种方法是：让一组红、蓝、绿的滤光片在传感器上方旋转。传感器快速、连续记录三幅独立图像。这种方法能够提供每个像素位置上全部三种颜色的信息。不过，这三幅图像不是准确地在同一时刻拍摄的；所以，在这三次读取数据期间，相机和图像目标必须保持不动。对于抓拍和手持相机而言，这个要求是不现实的。

一种更加经济、实用的记录单幅图像三基色的方法是：在每个光位上方永久放置一个滤光片，将传感器分为红、蓝、绿三组；这样，就可以从传感器的邻域获取足够信息，并对该处的真实色彩做出准确估计。这种利用附近传感器得到的其他颜色像素、对本地像素做出合理猜测的处理方法称为“插值”。

文件格式

TIFF 和 JPEG 是数码相机使用的两种主要文件格式。TIFF 属于未压缩格式，而 JPEG 属于压缩格式。大多数相机采用 JPEG 文件格式存储照片，有时也会允许对画质等级进行设置(如中或高)。下表将给出不同照片尺寸下图像文件的大致大小。(表略)

购买数码相机的注意事项

1. 确保相机拥有所需的分辨率。如果用电子邮件将拍摄的照片发给朋友，那么就不需要比 640×480 更高的分辨率。相反，如果想要将照片放大打印的话，那么就需要一架 200 万或 300 万像素的相机。

2. 确保相机有足够的存储器。再也没有比取景器中出现一幅很棒画面时，却出现“胶卷用完了”更扫兴的事了。数码相机的“胶片”就是闪存、“迷你盘”(minidisk)等设备。多数相机允许从相机下载图片，这样就可以拍摄更多照片；但如果去度假一周，那么就可能离开计算机而无法下载。因此，在购买相机时，应选择足够的额外存储器，这样在需要时就不会没的用。如今，Compact Flash 卡上有 1GB 空间，这个存储量肯定能够满足长途旅行的全部需要。

3. 确保镜头能够处理自己打算拍摄的画面。如果镜头不合适，就很难拍摄出最好的相片。例如：如果画面中的细节非常重要，那么就需要一个光学变焦范围大的镜头。在购买相机之前，一定要试一下相机上的镜头系统。数码相机的镜头种类很多，因此一定要到各处看一看。

4. 区分“数字变焦”和“光学变焦”。许多相机广告中说什么“100 倍变焦”，这往往是一种误导。因为其变焦功能只有一部分在镜头。变焦镜头唯一重要的部分是在“光学”部分，即由玻璃镜头组成的那部分。这才是提高图像质量的“变焦”。任何形式的“数字变焦”都是在相机之外自己可以完成的。如果使用相机软件在照片内切出一小块，然后将其放大，这就是在做“数字变焦”做的事。在多数情形下，数字变焦只会使图像变模糊。

5. 区分“实际分辨率”和“插值分辨率”。许多相机在广告中宣称有 1000×600 像素的分辨率和 1200×800 的插值分辨率。和数字变焦一样，插值分辨率是迷惑人的。借助相机软件，你自己也可以做这个事。插值无非就是使图像变得更大些、模糊些。

6. 看看电池能用多长时间。许多数码相机很费电——必须同时为图像传感器、液晶显示屏和微处理器供电，也许还有闪光灯！看看相机中的电池究竟能用多久。看看相机是否允许在紧急情况下使用标准的碱性电池。如果打算长时间使用相机的话，就应考虑另外购买一块电池。

第 24 课　电视视频信号

标准电视信号的出现已经有 50 多年历史了，它依然是最常见的图像传输方式之一。图 24.1 是在示波器上看到的电视信号。这个信号被称为“复合视频信号”，即在图像信号中加入了“垂直同步脉冲”和“水平同步脉冲”。这些脉冲用于电视接收机中的“同步垂直电路”和“水平偏转电路”中，用来配合正在显示的视频信号。标准视频信号在一秒钟内包含 30 幅完整图像，每一幅图像通常称作一“帧”(frame)。视频工程师会说每帧包含 525 线。“线”(line)也是电视术语，而程序员称之为“行”(line)。525 这个数字有点名不符实，因为只有 480～486 线包含视频信息；其余 29～45 线保留用于视频信号同步。

标准电视信号采用交错形式来减少画面的闪烁。也就是说，首先发送一帧画面全部的奇数线，然后发送偶数线。奇数线组被称作“奇数场”，而偶数线组被称为“偶数场”。每一帧由两场组成，因此视频信号每秒发送 60 场。每一场都是从一列持续 1.3 ms 的复杂垂直同步脉冲开始的。其后就是视频信号(偶数线或者奇数线)。每条线持续 63.5 ms，包括 10.2 ms 的水平同步脉冲，将一条线和另一条线分开。在每条线内部，模拟电压对应着图像的灰度，越亮的值表示沿黑色方向离开同步脉冲越远。因此，同步脉冲就在黑色范围之外。用视频术语讲，就是同步脉冲的电压值比黑色电压值还低。

在视频信号中，用于 A/D 转换的硬件称作“帧接收器”(frame grabber)。“帧接收器”通常是一张电子卡，它插入计算机、并通过同轴电缆连接到摄像机上。根据软件指令，“帧接收器”等候由垂直同步脉冲指示的下一帧的开始。在接下来的两场中，每条视频信号线要被采样多次，采样率的典型值为 512 样本/线、640 样本/线或 720 样本/线和 8 位/样本。这些样本就被存入存储器而成为数字图像中的某一行数据。

这种数字图像的采集方式导致在垂直方向和水平方向上存在显著区别。数字图像中的一行对应着视频信号中的一条线，也对应着 CCD 中的一行“势阱”(well)。而列就没有这么简单的对应关系。在 CCD 中，每一行包含约 400～800 个势阱(列)，具体数目是由所用的特定器件决定的。从 CCD 读取一行势阱时，视频信号线经过滤波后形成如图 24.1 所示的平滑模拟信号。换言之，视频信号并不取决于 CCD 有多少列。水平方向分辨率是受模拟信号可变速度限制的。在通常情况下，彩色电视信号设置为 3.2 MHz，上升时间约为 100 ns，即 53.2 ms 视频线的五百分之一。

视频信号在“帧接收器”中数字化之后，它就被转换回到列。然而，数字化图像中的列和 CCD 中的列是不相关的。数字化图像中列的数目仅由“帧接收器”对每条视频线的采样次数决定。例如，某个 CCD 每行可能包含 800 个势阱，而数字化图像每行可能仅含 512 个像素(即 512 列)。

此外，还有一个原因使得数字化图像中的列数很重要。标准电视图像的宽高比为 4∶3，即宽度稍大于高度。而电影图像的宽高比更宽(25∶9)。科学用 CCD 的宽高比通常为 1∶1，即一个完美的正方形。不论是哪种情形，CCD 的宽高比已经由电极的放置情况决定了，不能改变了。然而，在视频监视器上显示数字化图像或者对图像进行硬拷贝时，这就成了一个问题。假如不能正确重现宽高比，图像看上去在水平方向上或者在垂直方向上受到了挤压。

这里描述的 525 线视频信号称为 NTSC(国家电视系统委员会)制式。NTSC 制是 1954 年定义的标准。NTSC 系统用于美国和日本。在欧洲，存在两个相似的标准——PAL 制(逐行倒相制)和 SECAM 制(顺序与存储彩色电视系统)。PAL 制和 SECAM 制的基本概念是相同的，只是具体数字不同。PAL 制和 SECAM 制都采用每秒 25 个交错帧，每帧 625 线。和 NTSC 制一样，有些线在垂直同步时出现，使得约 576 线携带图像信息。其他不易理解的区别是和色彩、声音添加到信号的方式有关的。

传送彩色电视最直接的方法可能就是使用三路独立的模拟信号，分别代表人眼可以检测到的红、绿、蓝三种颜色。但令人遗憾的是，电视在其历史发展过程中并没有采用这种简单的机制。人们要求彩色电视信号做到这一点：在不做任何修改的前提下，黑白电视机能够继续使用。这是通过保留亮度信号，并增加一路独立的用于传输色彩信息的信号来实现的。在视频术语中，“亮度”被称作“辉度信号”，而“色彩”被称作“色度信号”。3.58 MHz载波上的色度信号被叠加到黑白视频信号上。声音信号以相同的方式加到黑白视频信号上，声音信号的载波是 4.5 MHz。电视接收机将这三个信号分离、分别进行处理，然后在终端显示器中再将其合成。

第 25 课　选择合适的微处理器内核

目前，在市场上广泛采用的 32 位微处理器内核大致有 7 种：摩托罗拉 680x0、英特尔 x86、PowerPC、MIPS、SuperH 和 ARM。当然，还存在着很多应用不太普遍的或专有性不很强的微处理器结构。其中，许多结构是和激光打印机、数字视盘(DVD)播放机等特定应用联系在一起的。

680x0 的使用量在下降，它也可能快走到头了。目前，680x0 主要用在 PalmOS 设备中。PalmOS“个人数字助理”(PDA)正在转向使用 ARM 器件，甚至摩托罗拉也在其旗舰 PDA 产品中采用了 ARM 内核微处理器。

基于高端 x86 系列微处理器的结构有如下明显优势：

- 可以使用几乎任何一种 PC 兼容的操作系统和免费的软件开发工具。
- 操作系统的安装非常简单。在多数情况下，“自动安装器”将检测系统的硬件构

成，并自动安装合适的内核、驱动程序等。这一点和嵌入式系统的规矩是不同的——（嵌入式系统）用户需要查看电路板、自己完成硬件配置的设置、（很可能要使用交叉编译器）生成系统内核和外部硬件的驱动程序。

• 对于几乎任何一种能想到的功能，x86 微处理器可以很容易地和数以千计的周边组件进行接口。这些组件是为消费市场生产的，巨大的生产量和激烈的价格竞争致使其价格便宜、容易获得。

• 所有你想添加到系统中的硬件的驱动程序几乎都有（大多数现成的操作系统中都包含这些程序）。

• 可以得到多种不同形式的、包含多种可能的外设组合的高集成度主板。

• 因缺乏支撑部件或因用户需求变化而进行的向大致相同硬件平台的移植相对要容易些。在许多情况下，这种移植仅仅需要对操作系统进行重新编译和安装，并为复制准备好一个新的主磁盘映像。

x86 系列器件的优点已经叙述过了，而其不足之处也必须指出。

• 和同样性能的 RISC 处理器比较，x86 器件非常昂贵。这就会影响到设备的商业化能力。

• 因为能够称得上是真正“片上系统”的 x86 器件比较少，所以在处理器之外很可能还需要一些外部硬件。为了获得某种特殊功能，往往需要添加一个多功能部件——因为得不到所需的单功能分立器件。这就提高了系统的复杂性和材料成本。

• x86 器件在功耗、散热、尺寸方面都存在明显的劣势。

• 现代 x86 器件及其支撑芯片都是高密度封装形式的高速器件。实际上，不可能以这种器件为基础手工设计样机；除非您想在设备上投入数万美元，但也不得不将其中某些装配工作承包出去。

• PC 外设集成电路的生产周期往往很短，12～18 个月是很平常的；因此，供货可能会成为一个问题。

• 用于裸机冷启动的代码往往很复杂，因为它必须对若干层（主板 BIOS、扩展卡 BIOS 及操作系统的各层）进行替换。CPU 结构也很复杂。

• 需要指出：在商用 x86 单板计算机上，基于 JTEG 的硬件调试系统或者其他硬件调试系统是不存在的。

当设备数量很少时或者当样机需要很多硬件功能而不想花费时间进行硬件调试时，建议选用 x86 作为硬件平台。当你既向早期用户供货、又在开发价格便宜的第二轮定制硬件设计的时候，x86 也是不错的选择。虽然还有其他适合选用 x86 的情况，但上述情况是主要的两种。

我们接着谈 RISC 平台。对于许多应用而言，MIPS、SuperH 和 PowerPC 都是很好的选择。其中，SuperH 是一个庞大的、包含各种各样有用器件的微处理器系列，而 MIPS 则是在第三方专用集成电路和专用标准产品中获得广泛授权的微处理器内核。PowerPC

主要在高性能应用中使用。对所有这些器件在不同的工程项目中的表现进行评价后，我发现用很少的资金去开发这些微处理器都是很困难的——评估硬件通常很贵，而且那些表现不出将会大量订货的购买者也不会立即得到这些器件。然而，在可预计的未来，所有这些处理器内核很可能都会存在下去，并且也会得到很好的技术支持。因此，只要你能够得到开发系统和器件，这些器件都是可行的选择。至少对于 SuperH 和 MIPS，设计基于该类处理器样机最省钱的途径是对现有某种硬件(如个人数字助理)进行重新定位。对于 PowerPC，我建议购买一块含有感兴趣芯片的单板工控计算机，但需要注意这样做也许很昂贵。PowerPC 不具有同 x86 兼容板一样的大规模市场价格，一块 PowerPC 单板机的价格是可比 x86 板价格的 2～3 倍。

除非某些 Intel 芯片的特点适合您的应用，我主要推荐的 32 位嵌入式平台是 ARM。该结构具有许多重要的优点(当然，这些优点中的一些也适用于上述其他 RISC 平台)。

- ARM 是一个成熟的、为人熟知的结构。它有翔实可靠的工程应用历史，并且做过许多改进。目前大量的 ARM 授权和大批 ARM 器件的出货可以确保其未来的供货能力。
- ARM 内核比较小，而且具有卓越的“功耗－性能”特性。
- 芯片设计者可以改变许多特性(协处理器、外部总线宽度、存储器管理单元和高速缓存大小等)。这意味着总能够找到一款满足几乎任何一种“性能/尺寸/功率”要求的 ARM 内核。
- 市场上存在大量价格诱人的、集成外设丰富的标准器件、定制器件和半定制器件。
- ARM 不仅提供了处理器内核的参考设计，而且提供了许多不同外设的参考设计；因此，在外设控制方面，不同的 ARM 实现方案(甚至是来自不同厂商的 ARM 实现方案)往往具有相似性。举个小例子，通常可以不怎么费力地就把一个用于串口数据发送的代码从一个 ARM 器件移植到另一个 ARM 器件。
- 部分由于上述因素，业界存在着大量用于 ARM 内核的免费 IP。

人们常说的一句话“ARM 就是 32 位的 8051”，就是在说 ARM 是一种“人人皆知、处处在用”的通用 32 位微处理器。这绝不是夸张，ARM 在嵌入式应用领域中的地位就如同 x86 在台式 PC 领域中的地位一样。

第 26 课　嵌入式系统设计语言

嵌入式系统是一种“看似不是计算机”的计算机，它能低成本、高效率地完成一组任务。典型的嵌入式系统可能需要完成通信、信号处理和用户接口的任务。这些任务需要

解决的问题不同；所以，能够解决所有嵌入式应用问题的通用语言是很难编写、分析和编译的。开发人员采用了多种不同的语言，每一种语言都有最适合的应用领域。

硬件语言

Verilog 和 VHDL 是最常用的硬件描述和建模语言。每一种“硬件描述语言”(HDL)都采用了(为高效仿真而忽略设计空闲部分的)离散事件语法对系统进行建模。这两种语言都使用了具有结构化层次的描述系统——将一个系统分成若干模块。系统模块包含了“基本元件”(primitive)实例、其他模块和并发进程。此外，每种 HDL 将这些连接都明确罗列出来。

Verilog 提供了更多的适于硬件仿真的基本元件。而 VHDL 的基本元件是像 a＝b＋c 这样的赋值表达式或程序代码。Verilog 增加了晶体管和逻辑门作为基本元件，而且允许用户使用真值表定义新的基本元件。

这两种语言都允许使用编写程序的方法来描述并发进程。在驱动事件唤醒之前，这些进程处于休眠状态；进程被唤醒后，进行变量读/写，然后终止。进程可能等待一段时间(如 Verilog 中的♯10，VHDL 中的 wait for 10 ns)，也可能等待某个值发生变化(@(a or b)，wait on a，b)，或者等待某个事件的发生(@(posedge clk)，wait on clk until clk ＝'1')。

VHDL 的通信更加规范和灵活。Verilog 通过导线或寄存器进行通信——共享存储器位置会导致出现竞争状态。VHDL 信号量的行为类似导线，但其判决函数可由用户定义。除被声明为共享变量外，VHDL 变量只在进程内部有效。

Verilog 的类型系统使用四值向量对硬件建模，用数组对存储器建模。VHDL 不含四值向量，但其类型系统允许用户添加数据类型。此外，用户还可以定义 C 结构体这样的复合数据类型。

总的来说，Verilog 是一种更简洁、更直接面向数字集成电路仿真的语言。而 VHDL 是一种更庞大、更详尽、能处理更多类型仿真和建模任务的语言。

软件语言

软件语言描述的是微处理器要执行的指令序列。大多数语言列出了要执行的指令，这些指令按照和存储器(存储器是存储数据的阵列)通信的顺序执行。

一般而言，每条机器指令只能做“两数相加”这么简单的操作。因此，高级语言的设计目标是能够简明、直观地指定指令。算术表达式是一种典型指令——用机器代码对 a^2+bx+c 这样的表达式进行编码虽然直接，但很冗长，编译器会完成得很好。C 语言提供了这种表达式，而且还提供了循环、条件等控制流结构和递归函数。此外，C＋＋语言提供了类的方法来创建新的数据类型、用于多态代码的模板、用于处理错误的异常机制和共用数据结构的标准库。Java 是更高级的语言，它提供了自动回收功能、线程和用来进行线程同步的监控程序。

汇编语言

汇编语言程序就是一张符号化的、人类可读形式的处理器指令列表。每条指令都包含一种操作(如加法)和一些操作数。例如,add r5, r2, r4 可以表示将寄存器 r2 和 r4 的内容相加的结果写入 r5。汇编语言依序执行这样的算术指令,而跳转指令可以通过改变程序计数器的值来执行条件语句和循环语句。程序计数器中存放着正在执行的指令地址。

某种微处理器的汇编语言是由操作码、寻址模式、寄存器和存储器来定义的。操作码用于区分不同的操作。例如,是加法还是条件转移。寻址模式定义了数据聚集、存储(例如,来自寄存器还是来自特定存储器位置)的方式。寄存器可被视作速度快且易于访问的小容量存储器。

C 语言

一个 C 程序中包含着由算术表达式构成的循环结构或条件结构的函数。C 程序指令按照顺序执行。但是,循环、条件等控制流结构会影响指令的执行顺序。当表达式中出现了函数调用,程序就将"控制权"交给被调用函数。该函数一直运行到输出结果,控制权返回并继续计算调用函数表达式的值。

根据微处理器能够直接操作的数据,C 语言可以推出自己的数据类型。这些数据类型包括有符号整数和无符号整数(从字节到字)、浮点数和指针。还可进一步将这些数据类型组合为数组和结构体("结构体"是由一些带名称字段组成)。

C 程序使用三种类型的存储器,在程序编译阶段,程序为全局数据分配空间。"堆栈"(stack)存储自动变量,程序在函数调用和返回数值的时候分配和释放这些变量。"堆"(heap)为程序提供了可以按照任意顺序重新分配的、大小任意的存储区域。尽管 C 语言是一种国际标准,许多人还是参考 K&R 写的书,因为 Rithie 设计了这种语言。

C++

为了适于开发大型程序,C++使用结构化机制对 C 语言进行了扩展。这些机制包括用户定义的数据类型、不同类型代码重用的方法、群对象命名规则、避免(程序片段进行汇编时出现的)名字冲突的命名规则和错误异常处理机制。C++标准库包含了数组、树、字符串等高效多态数据类型,编译器会为这些多态数据类型生成定制的实现。

"类"(class)通过指定其表现形式及可以访问和修改的操作,定义了一种新的数据类型。"类"可以用"继承性"对已有类进行扩展和修改。例如,一个"矩形类"可以将长度字段、宽度字段和一种计算面积的方法加到一个"形状类"中去。

"模板"(template)是一种能和多种数据类型工作的函数或类。编译器为每一个不同的模板生成定制的代码。

Java

Sun 公司的 Java 语言集合了 C++,但不兼容 C++。和 C++一样,Java 语言是面向对象的,它提供类和继承性。Java 是比 C++更高级的语言。因为 Java 使用了对象引

用、数组、字符串，而不使用指针。Java 的自动回收功能将程序员从存储器管理的工作中解脱出来。

Java 提供并发线程。创建一个线程需要对“线程”(thread)类进行扩展、为这些对象创建实例、调用它们的启动方法(以启动一个新的执行对象运行方法的控制线程)。

方法(或模块)的同步使用一种对象锁来解决(两个或更多进程试图同时访问同一个对象出现的)竞争问题。试图获得另一个线程拥有的锁的线程将停止运行，直至该锁被释放。使用这个特点，可以将特定对象的独享访问权授予某个线程。

实时操作系统(RTOS)

许多嵌入式系统使用“实时操作系统”(RTOS)来模拟单片处理器的并发性。RTOS 管理多个正在运行的进程，每个进程都是用 C 语言这样的顺序执行语言编写的。进程完成系统的运算，而 RTOS 对它们进行调度。为了满足“时限”(deadline)要求，调度机制将决定哪个进程何时、以何种顺序执行。

多数 RTOS 使用固定优先级的先占调度机制；在系统设计之时，每个进程就被赋予某个特定优先级(即一个小整数)。在任意时刻，RTOS 都运行具有最高优先级的进程；在运行一段时间之后，该进程就挂起、并等待更多数据的到来。通常，RTOS 使用“率单调分析”来分配优先级，最高优先级被分配给那些需要频繁满足“时限”的进程。

第 27 课　选择实时操作系统

工程师们常用“实时”(real-time)来描述这种计算问题：“应答迟滞”和“应答错误”一样糟糕。这一类问题对“时限”(deadline)是有限制的，嵌入式系统常常在这种约束条件下运行。例如，假如控制防抱死刹车装置的嵌入式软件的某个“时限”没有得到满足，那么车祸就可能发生。所以，对于实时嵌入式系统的设计人员，尽可能了解其软、硬件的行为和性能是至关重要的。

实时系统设计人员要花费大量时间来考虑“最坏情况”(worst-case)下的系统性能。他们必须不断地问自己这样的问题：在“最坏情况”下，“驾驶员踩刹车踏板”和“中断信号抵达处理器”这两个事件的时间差是多少？在“最坏情况”下，“中断反应时间”(interrupt latency)是多少？在“最坏情况”下，软件需要多长时间才能触发刹车系统？仅仅进行“平均情况分析”或“期望情况分析”是不够的。今天，大多数商用嵌入式操作系统在设计时就都将实时系统考虑在内了。在理想情况下，这些嵌入式操作系统的最坏性能是明确的、可信的。

要想获得“实时操作系统”的名称，这个操作系统应该是“确定的”(deterministic)，并

要确保“最坏情况下”的“中断反应时间”和“任务切换时间”(context-switching time)。当给定这些特征以及系统中的任务和中断的“相对优先级”(relative priority)时,就有可能使用“率单调分析法”(rate monitonic analysis)对软件的最坏情况性能进行分析了。

如果每个“系统调用”(system call)在最坏情况下的执行时间都是可计算的,那么这个操作系统就是“确定的”。如果一个操作系统提供商能够认真看待其产品的实时性能的话,那么一般都会通过数据手册公布每个系统调用所需时钟周期数的最小值、平均值和最大值。对于不同处理器而言,这些数值可能不尽相同;但是,假如算法在某个处理器上运行时是“确定的”,可以合理地推断这个算法在任何其他处理器上运行都是确定的。(当然,实际情况可能会和这个推断不同。)

“中断反应时间”是指从中断信号到达处理器到“中断服务程序”(ISR)开始运行之间的全部时间。当中断出现时,在执行 ISR 之前,处理器必须完成以下几步。首先,处理器必须停止正在执行的指令。然后,必须辨认中断类型。这是由硬件完成的,不会减慢或挂起正在运行的任务。最后,只有当中断使能、CPU 环境设置被保存之后,相关的 ISR 才会启动执行。

当然,假如中断会被禁用(例如,在系统调用中),那么“最坏情况”下的“中断反应时间”将增加,增加量是中断的最长关闭时间。每个操作系统都会在一些地方禁用中断一段时间;因此,就需要知道系统的需求是什么。一个实时工程项目可能需要“中断响应时间”在 1 μs 之内,而另一个实时工程项目可能仅仅需要在 100 μs 之内。

操作系统的第三个实时特性是“任务切换”所需的时间。这个特性之所以重要是由于它反映了整个系统的开销。例如,假定任何一个任务的平均执行时间是 100 μs,而任务切换时间也是 100 μs。这样的话,整整一半的处理器时间都用在任务切换例程上了!

这里也没有什么特定的数据。实际时间通常是由处理器决定的,因为这取决于应予保护的寄存器数量和在何处进行保护。一定要从操作系统提供商那里得到这些数据;这样,在即将购买前就不会出现任何意外。

选择过程

考虑到当今工程的时间成本,花几千美元买一个商用实时操作系统是合算的。对于大多数工程项目和资金预算,是可以买到合适的操作系统的。商用操作系统在功能、性能和价格上是一致的。低端商用操作系统仅仅提供基本的“先占式调度器”(preemptive scheduler)和几个关键的系统调用。这种操作系统往往价格低廉、源代码无法修改、无需支付版权费。高端操作系统往往包含除基本调度器之外的许多功能。这种操作系统都相当昂贵,尽管“启动成本”(startup cost)在 10 000～ 50 000 美元之间,每个 ROM 复制都要付版权费。不过,这个价格往往包含了免费技术支持、技术培训和一套集成开发工具。在这两种极端情况之间,还有一些操作系统价格适中,但不包含源代码,技术支持也会额外收费。大多数商用操作系统属于这一类。

操作系统种类如此繁多，特点各不相同，很难决定哪个操作系统最适合你的工程项目。首先，要考虑你对处理器、实时性能和经费预算的要求。你无法改变这些指标，因此使用这些指标可以将可选范围缩小一些。然后，和这些操作系统的提供商联系，获取更加详尽的技术信息。

此时，许多人就会根据操作系统和已选交叉编译器、调试器和其他开发工具之间的兼容性做出决定了。不过，你也应该搞清楚：对于你的工程项目，哪些附加功能是很重要的？不论你购买什么附加功能，基本内核大体都是一样的。区别很可能在于：处理器的支持情况、最大和最小存储器要求、有无附加软件模块（如网络协议栈和驱动程序）以及是否和第三方开发工具兼容。

选择商业操作系统最具说服力的理由是可以利用经过充分验证的内核，这要比使用自己开发内核的可靠性高得多。因此，在选择操作系统提供商时，考察其开发经验如何是很重要的一点。

第 28 课　信号源

完整的测量系统

在进行电子测量的时候，人们首先想到的很可能是一件“采集仪器”——通常是示波器或逻辑分析仪。只有能够获得某种类型的信号时，这些工具才能进行测量。但是，在许多情形下，这样的信号是不存在的，除非外部提供。

举例来说，一个应力测量放大器不会产生信号；它仅仅能够将传感器接收到的信号放大。同样，数据总线上的“多路复用器”也不会产生信号；它为来自计数器、寄存器和其他元件的信号指引方向。但是，在将放大器或多路复用器连接到提供信号的线路之前，不可避免地要进行必要的测试。为了使用采集工具测量这些器件的行为，必须要在其输入端提供激励信号。

举另外一个例子。为了确保新硬件在整个工作范围以及该范围之外满足性能设计要求，工程师必须对即将完成的设计进行特性评价。这就是所谓的“边际测试或极限测试”。这件测量工作就需要一个“完整的”解决方案——一个既能进行测量、又能产生信号的方案。

用于数字设计特性评价的测量仪器和用于模拟/混合信号设计的测量仪器是不一样的，但两者一定都包含激励信号产生仪器和信号采集仪器。信号源（或信号发生器）和采集仪器是构成一个完整测量解决方案的两大基本组成部分。

什么是信号源?

信号源是用于硬件设计、调试、评价工程中的几乎任何一种测量配置的基础。信号源是一种关键的工程工具。对技术人员而言,它是必备的故障诊治助手。它可以产生用来代替汽车点火脉冲、心脏起搏器或导弹陀螺仪的输出信号。在电子测试仪器中,也许信号源是仅次于(无处不在的)“数字多用表”(DMM)的最通用的仪器。

当对元器件、功能模块和子系统进行验证时,电路中很可能需要某种类型的“交流”(AC)激励源,除非你只使用纯“直流”(DC)电路。信号源输出的波形模拟了来自外部世界的信号(如传感器输出信号)。同样,信号源还可用于替代电路设计中尚未完成部分所产生的波形。

有趣的是,信号源不仅提供一个“理想”的波形。它经常在输出波形中添加已知的、数量和类型均可重复的“畸变”(误差)。由于往往不可能使用电路准确地在所需时间、位置产生可预测畸变,所以信号源的这个功能是极为有用的。“被测单元”(UUT)对畸变信号的响应反映了该单元在正常环境下处理问题的能力。

激励信号可以采取低畸变正弦波、逻辑脉冲、高频无线载波、移动电话发射信号等多种形式。传统上,这些波形是由独立的专用信号源产生的——从超纯净音频正弦波发生器到 GHz 射频信号发生器。尽管市场上存在着多种商业解决方案,但是为了应用到手头的项目中,用户常常要对信号源进行定制设计或修改。设计一个仪器级的信号发生器是非常困难的,而设计辅助设备耗费的时间也将人们的宝贵精力从项目中分散开。

幸运的是,数字采样技术和信号处理技术已经带给人们一种使用一种仪器就可以解决几乎所有类型信号产生需求的方案——任意信号发生器。

数字信号源的类型

数字信号发生器囊括了全部信号产生需求,它可分为任意波形发生器(AWG)、任意函数发生器(AFG)和数据/模式发生器(DG)三大类。每一类都有其独特的优点:

- AWG:可以产生你想象到的任何波形——不论是用于测试磁盘驱动器的洛伦兹脉冲整形数据流,还是测试 GSM 手机或 CDMA 手机的复杂调制射频信号。可以使用多种方法(从数学公式到波形绘制)创建所需的输出信号。

- AFG:提供的波形种类一般比较少,但它具备卓越的稳定性,而且能快速响应频率的改变。假如 UUT 需要典型“正弦波和方波”和在两个频率间的瞬时切换能力,那么 AFG 就是正确的选择。AFG 的另外一个优点是价格低,这使其在无需 AWG 灵活性的应用中非常具有吸引力。

- DG:能够满足需要具有特定内容和时序特征的长时间连续二进制数据流的数字器件对特殊激励源的需求。

信号产生技术

使用信号源生成波形的方法有许多种。选择何种方法取决于“被测设备”(DUT)的已知信息和该设备对输入信号的要求——是否需要添加畸变、误差信号及其他变量。现代高性能信号源提供至少三种波形产生的方法：

• 模拟：以波形的定义(往往来自模拟器或者波形库)为基础，创建一个事件或一个事件序列。

• 复制：记录示波器上存在的波形并将其发送到信号源进行重建(图 28.1)。

• 替代：创建或者修改一个已定义信号，用它替代一个本应由电路产生、但却无法得到的信号(图 28.1)。

信号源的主要用途

信号源有几百种用途。在电子测量领域中，信号源的应用分为以下三类：功能验证、特征量提取和极限性能测试。一些具有代表性的应用如下：

• 功能验证——数字调制分析：开发新型发射、接收硬件的无线设备设计商必须对基带 I、Q 信号(包括有损和无损)进行仿真，以确保它们研发的设备能够符合刚出现的和已有的无线通信标准。某些高性能 AWG 能够提供数据率高达 1 Gbps 的两路独立的低畸变、高精度信号——一路产生 I 信号，一路产生 Q 信号。

• 特征量提取——测试“数模转换器”(DAC)和“模数转换器”(ADC)：为了确定线性度、单调性和畸变的范围，新开发出的 DAC 和 ADC 必须进行全面测试。AWG 的现有技术水平能够同时产生同相模拟和数字信号，并驱动数据率高达 1Gbps 的器件。

• 极限性能测试——通信接收机的极限性能：使用串行数据流结构(普遍用于数据通信总线和磁盘驱动放大器中)的工程师需要用“损害”(抖动和时序违例)来测试器件的极限性能。高级的信号源提供了高效的内置式“抖动编辑和生成工具”，这省去了工程师大量的、花在计算上的时间。这些仪器能够将关键的“信号边沿”(signal edge)移动0.3 ps。

信号源硬件结构：任意波形发生器

从根本上说，AWG 是一种复杂的“重放”系统——波形的产生是以(描述交流信号不断变化电压的)存储数据为基础。AWG 的结构框图看似很简单。

为了理解 AWG，先要掌握“数字采样”这个概念。“数字采样”是指使用(沿波形变化测量到的)一组电压样本(数据点)来定义一个信号。这些样本既可能是通过使用示波器等仪器对波形进行实际测量得到的，也可能是利用图形化技术或数学技术得到的。

图 28.2a 画出了一组采样点。所有采样点的时间间隔都是相等的，尽管曲线使得这些间隔看上去不一样。在 AWG 中，采样值以二进制形式存储在快速 RAM 中。

通过读取存储器地址和输入数据到 DAC，可以利用存储信息随时重建信号（见图 28.2 b）。注意：AWG 的输出电路在样点间进行滤波，将这些点连接起来，产生纯净的、不间断的波形。UUT 看到的不是离散的点，而是连续的模拟波形。图 28.3 是 AWG 实现上述过程的简化框图。

在这里，还需要定义一些术语和说明影响信号重建保真度的条件。在这些术语和条件中，首要的是奈奎斯特定理。该定理指出：采样频率至少为被采样信号最高频率分量的两倍。例如，为了对 1 MHz 的信号进行采样，至少需要 2 MS/s 的频率进行样本采集。

尽管该定理通常作为信号采集（如示波器）的指南，但它和 AWG 的关系是显而易见的——为了能够真实再现期望信号的细节，存储波形必须拥有足够多的采样点。

AWG 按照指定范围内的任意频率获取并从存储器读出这些样点。如果一组存储样点满足奈奎斯特定理，而且描述了一个正弦波，那么 AWG 的输出将是一个正弦波形。当然，该仪器存在一个最高工作频率（采样率）。通常，这个频率用 MS/s 或 GS/s 表示。

今天，最快的 AWG 可以达到 2.6 GS/s。AWG 的其他硬件特点也同样重要，特别是垂直分辨率、存储深度和采样率。“垂直分辨率”（幅度分辨率）表示的是上述样点幅度测量的精度。这个精度和仪器 DAC 的二进制字长（以 bit 为单位）有关——位数越多，精度越高。尽管“越多越好”，高频 AWG 的精度（8 位或者 10 位）往往比通用 AWG（12 位或 14 位）要低。

在仪器满幅电压范围内，10 位精度的 AWG 提供了 1024 个采样电平。例如，若一个 10 位的 AWG 电压范围为 2Vp-p（峰-峰值），则每个样本代表着约 2 mV 的量阶——假定不受结构中其他因素的限制，这就是该仪器提供的最小增量。

“存储深度”在仪器的灵活性方面具有关键作用。大存储深度提供两个好处：

1）可以存储更多周期的期望波形。这很有用，因为减少了“端点”数。“端点”是指波形的最后存储器位置，在这个位置之后 AWG 必须折回到存储器的开始位置，才能连续产生输出信号。由于转换时发生的误差是不可避免的，所以希望减少端点数量。

2）可以存储更多的波形细节。在复杂波形的脉冲边沿处和波形过渡处存在着高频信息。在这些快变区域，很难采用和简单的、可预测的正弦波形一样的插值方法。为了真实再现一个复杂信号，现有的波形存储容量必须用于存储更多的过渡波形和振荡波形，而不是更多的信号周期。

今天的 AWG 技术水平可以提供 8 M 的存储深度。某些顶级型号的 AWG 同时提供大存储深度和高采样率。这些仪器可以存储和重现复杂的射频波形，甚至包括用于网络设备物理层测试的伪随机数据流。同样，这些 AWG 能够产生非常短暂的数字脉冲和暂态波形。

第 29 课　示波器

自然界的运动形式就是正弦波，不论是海浪、地震、声爆、爆炸、空气中传播的声音，还是运动物体的本征频率都是如此。整个宇宙遍布着能量、振动粒子和其他人眼看不见的作用。甚至是光线——半波、半粒子——也有基波频率，这个频率在人看来就是颜色。传感器可将各种作用转化成为使用示波器可以测量和研究的电信号。有了示波器，科学家、工程师和教师等人员就能够看到事件随时间变化的过程。

对于电子设备的设计、生产和维修人员而言，示波器是不可或缺的工具。在如今这个迅速变化的世界里，工程师需要最好的工具迅速、准确地解决测量问题。示波器就像工程师的眼睛一样，它是完成当今苛刻测量任务的关键。

示波器的用途不仅限于电子技术领域。通过使用适当的传感器，示波器可以测量各种现象。传感器（换能器）是一种能够受其他物理量（如声音、机械张力、压力、光和热）激励而产生电信号的器件。麦克风就是一种传感器，它将声音转换为电信号。

人人都可以使用示波器——从物理学家到电视维修技工。汽车工程师使用示波器测量发动机的振动，而医学研究人员使用示波器测量脑电波。示波器具有无数种可能的用途。

示波器类型

电子设备可分为模拟设备和数字设备两大类。模拟设备处理的是连续变化的电压，而数字设备处理的是代表电压样值的离散二进制数字。

示波器也有类似的分类——模拟示波器和数字示波器。对很多应用而言，二者都可胜任。然而，它们也有各自独特之处；对特定应用而言，这些独特之处可能使之更适用，也可能使之不很适用。数字示波器还可进一步分为"数字存储示波器"(DSO)、"数字荧光示波器"(DPO)和"采样示波器"三类。

模拟示波器

模拟示波器的基本工作原理是将被测信号电压直接加到电子束的垂直系统，控制着电子束在示波器荧光屏（通常是"阴极射线管"，CRT）上从左向右扫描。荧光屏背面涂有一层荧光物质，电子束撞击到的地方就会发光。当信号在显示屏水平移动时，信号电压就成比例地将电子束上、下偏转，从而在荧光屏上显示出信号波形。电子束撞击荧光屏上哪个位置越频繁，哪个位置就越亮。

CRT 限制了模拟示波器可显示的信号频率范围。当频率很低时，信号看上去就是一

个明亮的缓慢移动的亮点，很难让人辨别出波形的形状。当频率很高时，CRT 的显示速度又形成了限制。当信号频率超过 CRT 的显示速度时，信号就会变得太暗而无法看清。最快的模拟示波器能够显示大约 1 GHz 的信号。

当把示波器“探头”(probe)连接到电路中的时候，电压信号将通过“探头”抵达示波器的垂直系统。图 29.1 展示了模拟示波器显示被测信号的方法。根据垂直刻度(伏/格)的设置，电压信号可能经由衰减器而被衰减，也可能经由放大器而被放大。

然后，信号就直接抵达 CRT 的垂直偏转板。偏转板两端的电压会导致示波器荧光屏上出现一个移动的发光点。这个发光点是由电子束轰击 CRT 内壁的荧光物质而产生的。正电压会使发光点上移，而负电压会使发光点下移。信号也会抵达示波器的触发系统，从而启动或者触发一次水平扫描。“水平扫描”是指水平系统使发光点在荧光屏上移动。

触发“水平系统”能使水平“时基”驱动发光点在指定时间间隔内在荧光屏上从左到右移动。多次快速扫描会使发光点汇合为一条实线。在更高的速度上，发光点在荧屏上往返的速度可达 500 000 次/秒。

水平扫描和垂直偏转一起在荧光屏上勾勒出信号的形状。为了稳定一个重复信号，需要进行“触发”——它保证每次扫描重复信号时的起始点都一样。

此外，模拟示波器都有聚焦、亮度调节控制功能，能够产生清晰的显示。当需要“实时显示”(即刻显示)快变波形时，人们往往想起模拟示波器。模拟示波器的荧光显示器具备一种“亮度分级”特性——频繁出现的信号特征具有更亮的扫描线。有了这种“亮度分级”特性，仅通过查看扫描线的亮度级别就可以很容易地区分信号细节。

数字示波器

和模拟示波器相反，数字示波器采用“模数转换器”(ADC)将被测电压转换为数字信息。数字示波器采集到的波形是一组样本，并将这些样本保存，直至样本数达到足以描述波形的数量。之后，数字示波器将这些样本重新组合、在示波器显示屏上显示。数字示波器可分为“数字存储示波器”(DSO)、“数字荧光示波器”(DPO)和采样示波器。

数字技术意味着示波器可以稳定、清晰地显示示波器范围内的任何频率信号。对于重复性信号，数字示波器的带宽是示波器前端器件模拟带宽(一般是指 3 dB 带宽)的函数。对于只出现一次的信号和暂态信号(如脉冲信号和阶跃信号)，数字示波器的带宽可由示波器的采样率决定。

数字存储示波器

“数字存储示波器”(DSO)是一种常见的数字示波器。DSO 的显示取决于光栅显示屏，而不是荧光显示屏。DSO 能够捕捉、观测那些可能只出现一次的事件——暂态事件。由于波形信息是以一组二进制值的数字形式存储的，所以可以对这些信息进行分析、存档、打印和处理；这个过程既可以在示波器内部进行，也可以在计算机上进行。波形不必

是连续的；在信号消失之后，DSO 仍然可以显示波形。和模拟示波器不同的是，DSO 能够永久保存信号，并可以对波形进行各种处理。然而，DSO 一般没有实时亮度分级能力；因此，它无法反映信号亮度级别的变化。

数字荧光示波器

“数字荧光示波器”（DPO）采用了一种新型示波器结构。该结构使得 DPO 具备精确重建信号的独特的采集、显示能力。DSO 使用串行处理结构（见图 29.2）对信号进行采集、显示和分析，而 DPO 使用并行结构来实现这些功能（如图 29.3 所示）。DPO 结构使用专用、独特的 ASIC 硬件采集波形的映像；波形采集速率高，信号可视化程度就高。这就提高了观测到数字系统暂态事件（短脉冲、毛刺和转换错误）的概率。

数字采样示波器

在测量高频信号时，示波器在一次扫描中可能无法搜集足够多的样本。当需要准确记录那些频率成分远远超过示波器采样频率的信号时，数字采样示波器就是一种理想的工具。这种示波器能够测量的信号速度要比其他示波器能够测量的信号速度快一个数量级。对于重复信号而言，数字采样示波器的带宽和高速时序是其他示波器的 10 倍。连续等效时间采样示波器的带宽可达 50 GHz。

第 30 课　逻辑分析仪

逻辑分析仪的功能和示波器是不同的。最明显的区别是两者输入信号通道数不一样。一般而言，数字示波器的通道数可达四路，每个通道输入一路数字信号。有些复杂的系统设计需要几千路输入通道，也存在相应规模的逻辑分析仪。

逻辑分析仪测量、分析信号的方式也和示波器不一样。逻辑分析仪并不测量信号的模拟细节，而是检测信号的逻辑门限电平。当把逻辑分析仪连接到数字电路的时候，人们只关心信号的逻辑状态。逻辑分析仪只看两个逻辑电平：当输入信号电压超过门限电压时，输入信号就被认为是“高电平”或者“逻辑 1”；当输入信号电压低于门限电压时，输入信号就被认为是“低电平”或者“逻辑 0”。逻辑分析仪对输入信号进行采样的时候，它将根据信号电平和门限电平的相对大小存储一个“1”或者一个“0”。

逻辑分析仪的“波形时序显示”和数据手册上提供（或者模拟器产生的）的“时序图”很相似。所有信号都和时间相关，这样就可以观察到“建立－保持时间”、脉冲宽度、外来数据或丢失数据。除了通道数量不同之外，逻辑分析仪还提供了许多支持数字设计和验证的功能，其中包括：

- 复杂的触发机制：允许用户指定逻辑分析仪进行数据采集的条件
- 高密度探头和适配器：简化了逻辑分析仪和“被测系统”（SUT）之间的连接

• 分析能力:将采集到的数据转化为处理器指令,并将这些指令和源代码相关联

结构和操作

逻辑分析仪和数字信号相连,采集、分析数字信号。使用逻辑分析仪有如下四步:

1. 探测(将逻辑分析仪连接到 SUT 上)
2. 设置(设置时钟模式和触发模式)
3. 采集
4. 分析和显示

图 30.1 是逻辑分析仪的简图。每个方框代表了若干个软、硬件模块。方框上的编号和上述四个步骤相对应。

探头

逻辑分析仪可以一次采集大量信号,这一点就和示波器不同。探头连接到 SUT 上。探头的内部比较器将输入信号电压和门限电压 V_{th}进行比较,并判决信号的逻辑状态(1 或者 0)。门限电压值是由用户设置的,设置范围从 TTL 电平到 CMOS 电平、ECL 电平和用户定义电平。

探头能够采集高质量信号,而且对 SUT 的影响最小。探头为逻辑分析仪提供了一个高质量的信号通道,并将对 SUT 造成的电气负载降至最小,而且能够适配于各种电路板和器件。

逻辑分析仪探头的阻抗(电容、电阻和电感)会成为整个被测电路负载的一部分。所有探头都会表现出负载特性。逻辑分析仪探头引入到 SUT 中的负载应该尽量小,并为逻辑分析仪提供准确的信号。

探头的电容会将信号边沿的过渡过程变得缓慢。这一点为什么很重要呢?因为过渡缓慢的边沿穿越电路门限电平的时间要晚一些,这样就会引入时序误差。随着时钟频率的提高,这个问题就会变得很严重。在高速系统中,过量的探头电容可能会(潜在地)阻止被测系统工作!选择一个总电容尽可能小的探头往往是很关键的。

设置时钟模式和触发模式

逻辑分析仪可以从多引脚器件或者总线上采集数据。“采集速率”这个词是指输入信号的采样频率。其功能和示波器中的时基电路是一样的。数据采集有两种类型:

1)“异步采集”是对信号的时序信息进行采集。在这种模式下,逻辑分析仪的内部时钟用于数据采集。数据采样越快,测量精度就越高。在目标器件和逻辑分析仪的数据采集之间不存在固定的时序关系。这种采集模式主要用于观测 SUT 信号间的时序关系。

2)“同步采集”用于获取 SUT 的状态。SUT 的信号决定了样本点(何时、以什么频率采集数据)。作为同步采集时钟的信号可能是系统时钟、总线控制信号或者一个引发

SUT 状态发生改变的信号。在有效边沿对数据进行采样，这个数据反映了逻辑信号稳定时的 SUT 状态。当且仅当被选信号有效时，逻辑分析仪才进行采样。

数据采集：实时采集存储器

逻辑分析仪的探测、触发和定时系统将数据传送给实时采集存储器。这个存储器是逻辑分析仪的心脏——它既是所有 SUT 的采样数据的目的地，也是逻辑分析仪所有分析显示功能的数据来源。

逻辑分析仪的存储器能够按照仪器的采样率存储数据。这种存储器可想象为有一定宽度和一定深度的矩阵。

在一个触发事件出现或者用户告知停止前，逻辑分析仪一直在积累数据。这就是采集——具有很高时序准确性的多通道波形显示(可用于观测所有被采集信号间的相互作用)。

选择逻辑分析仪时，宽度和深度都是关键因素。下面几条提示会有助于确定通道数量和存储深度：

你需要记录、分析多少路信号？逻辑分析仪的通道数和希望记录信号数是直接对应的。数字系统总线的宽度有很多种，在总线检测的同时还需探测一些其他信号(时钟信号、使能信号等)。一定要考虑到你要同时记录的所有总线和信号。

你需要记录多长时间？这将决定逻辑分析仪的存储深度，而这一点对于异步采集尤为重要。对于给定的存储容量，随着采样率的提高，总的采集时间会下降。例如，当采样率为 1 ms 时，1 M 存储器存储的数据对应的时间跨度为 1 s。当采样时钟周期为 10 ns 时，同样的 1 M 存储器只能存储 10 ms 的数据。采集的样本多(采集的信号时间长)，捕获错误、(导致错误的)问题的机会就大。说到存储容量，当然是越多越好！

分析和显示

实时采集存储器中存储的数据可用于几种不同的显示和分析模式。信息一旦存到系统当中，就可以按照多种格式(从时序波形到和源代码指令相关的助记符)观察这些信息。

“波形显示”可以看到多个通道的详细情况，你可以看到所有记录信号之间的时间关系，这很像示波器的显示。“波形显示”常用于时序分析，它极适合完成如下工作：

- 诊断 SUT 硬件的时序问题
- 验证硬件工作是否正常(通过对记录结果和仿真器输出或数据手册时序图进行比较)
- 测量硬件时序特性
  —竞争状态
  —传输延迟
  —脉冲的有无

• 分析毛刺

“列表显示”可以提供“用户选择的”字母数字形式的状态信息。列表中的数值是从整个总线上记录的样值得来的，这些数值可以表示为十六进制格式，也可以表示为其他格式。“列表显示”的目的在于展示 SUT 的状态。“列表显示”展示给用户的信息流就是 SUT 看到的信息流——一组数据字。

状态数据的显示格式有好几种。“实时指令跟踪”将每个总线操作进行反汇编，准确判定总线读取的是哪条指令。它将对应的“助记符指令”及其地址显示在逻辑分析仪显示屏上。

还有一个附加显示，就是“源码调试显示”。通过将源码和过去的指令跟踪情况关联起来，它能使调试工作更有效率。有了“源码调试显示”，人们即刻就可以看到在一条指令执行时，到底发生了什么。

练习参考答案

第 1 单元

填空

（1）Component，All work that you do，indirectly，if not，amazing，devices，plays a significant role，aspects，determinant，While，dictate，reach its full potential，enable，In，such as，how reliably your system will run，dependable，to some extent，consumes，to，a great deal of，on，major，in turn

（2）solid state，electronic，non-volatile，high density，code and data，making，embedded，of，ability，backup，as soon as，available，From a developer's viewpoint，so that，offline，that，taken，and，Because of，erased，Rather than，repeated，that must be considered，have a damaging effect on，For any given flash device，to，or

英汉互译

1）现代社会的许多产品和服务都是建立在电气工程师和计算机科学家的工作基础之上。在过去几十年中，数字电子器件成本的大幅下降已经带来了计算机应用迅猛的增长。与此同时，人们对计算机科学认识的提高使得开发新的更强大、更复杂、更灵活的软件系统成为可能。

2）1971 年 11 月，英特尔公司的三位工程师发明了世界上首片商用微处理器 4004。以今天的标准看，4004 是比较简单的。它仅包含 2300 个晶体管，在 1 秒内可执行约 60 000次运算。今天，微处理器是最复杂的大量生产产品。它包含几百万个晶体管，1 秒中可完成数亿次运算。

3）微处理器是个人计算机的神经中枢。在这个小小的硅片上集成了数以百万的（数字）开关和（数据）通路，它们帮助计算机做出重要的决定和完成有意义的任务。但微处理器不是只为计算机而设计的。人们可以在电话、汽车等许多日常物品中发现处理器。

4）存储器是计算机和嵌入式系统中用来保存信息的部件。在嵌入式应用中需要存储信息，因此，设计中总含有某种类型的存储器。现在有许多种可以利用的存储器件，因此对于嵌入式用户存在多种选择。

5）存储器可以分为两大类：易失的和非易失的。易失存储器在系统掉电后就丢失了其（存储）内容。非易失存储器系统或者器件掉电之后不会丢失数据。

6）Modern electronic systems are increasingly digital in nature，exceedingly complex，and would be inconceivable without today's VLSI chip technology. Indeed，such systems are so complex that the principles of their design bear great similarities to the design principles of large software systems. Thus，computer science and electronic system design require similar backgrounds in many respects，and computer aids to design are essential in this ever-expanding domain of engineering.

7）A chip is a small piece of semiconducting material on which an integrated circuit is embedded. A typical chip is less than 1 square inches and can contain millions of transistors. There are different types of chips. For example，CPU chips contain an entire processing unit，whereas memory chips contain blank memory.

8）A microprocessor is a silicon chip that contains a CPU. There are three basic characteristics differentiate microprocessors：instruction set，bandwidth and clock speed.

9）There are several reasons to use Flash memory instead of a hard disk：Flash memory allows faster access. It is noiseless，lighter，and smaller in size and has no moving parts.

10）Memory is the part of a computer system that is used to run programs and store data. The main memory is built from Random Access Memory（RAM）chips. The amount of memory available determines the size of programs that can be run，and whether more than one program can run at a time. Main memory is temporary，and the data is lost when the computer is turned off. It is distinguished from more permanent internal Read Only Memory（ROM），and external storage. ROM contains the computer's essential programs In the general sense，Memory can be any device that can hold data in machine-readable format.

第 2 单元

填空

（1）at，while，in，In the field of telecommunication，modems and speech processing，channel selection，anti-aliasing low-pass filters，signal conditioning，filter，a liner phase shift，time delay，all-pass filters，of，such as，called，making，cases，active filters，an operational amplifier，in，resistors and capacitors

（2）passive components，as，by，from，ignorant of，analog circuitry，limited frequency ranges，to，radiated or conducted，as well，a high-pass filter

英汉互译

1）由于很难定义运算放大器的“动态范围”（DR），因此先给“数模转换器”（DAC）的动态范围下定义。DAC 的动态范围就是其最大输出电压和最小输出电压之比。

2）这个“动态范围”的定义也可用于运算放大器，其最大输出电压摆幅为 V_{OUTMAX}。输出电压摆幅定义为运算放大器能达到的最大输出电压（V_{OH}）减去最小输出电压（V_{OL}）。从运放集成电路数据手册中，可以很容易获得 V_{OH} 和 V_{OL}。

3）噪声是在一段时间内随机变化的，信号或噪声电平的瞬时值就不足以描述这种情形。信号和噪声要用其在较长时段内的平均值（均方根值，RMS）来描述。“信噪比”（SNR）用于度量噪声环境中的信号质量。SNR 是功率比，是在电路输出端得到的。我们感兴趣的 SNR 是电压比（因为阻抗是常数），是在运算放大器的输入端得到的。就是说，全部噪声（包括电阻噪声）电压必须在运算放大器的输入端以均方根值进行计算。

4）切比雪夫等波纹滤波器的频域滚降分布在整个通带范围内。其通带内波纹较多、而过渡区的滚降比较陡峭。因其在滤波过程中的 Q 值较高，切比雪夫等波纹滤波器的暂态响应和阶跃响应较差。

5）芯片封装就是芯片的外罩。有了它，才能把芯片插入（插座安装）或焊至（表面安装）印制电路板上。在外行看来，芯片安装设计似乎不太重要；事实上，芯片封装技术却是一个庞大、复杂的行业。为尺寸不断缩减的电路小片（裸片）提供越来越多的“输入/输出互连”能力是经常出现的问题。除此之外，封装和半导体电路的“缩微化”均有助于减小蜂窝电话和其他手持设备的尺寸。

6）In one respect, voltage is like water: you don't appreciate its value until your supply runs low. Low-voltage systems, defined here as a single power supply less than 5 V, teach us to appreciate voltage.

7）Ohm's law is stated as V＝IR, and it is fundamental to all electronics. Ohm's law can be applied to a single component, to any group of components, or to a complete circuit. When the current flowing through any portion of a circuit is known, the voltage dropped across that portion of the circuit is obtained by multiplying the current times the resistance.

8）Circuits are a mix of passive and active components. The components are arranged in a manner that enables them to perform some desired function. The resulting arrangement of components is called a circuit or sometimes a circuit configuration. The art portion of analog design is developing the circuit configuration.

9）Noise sets a limit on the information and signals that can be handled by a system. The ability of an amplifier, receiver, or other device to discern a signal is degraded by noise. Noise mixed with the incoming signal, noise generated by the op

amp, resistor noise, and power supply noise ultimately determine the size of the signal that can be recovered and measured.

10) A Butterworth (maximally flat) filter is the most commonly used general-purpose filter. It has a monotonic passband with the attenuation increasing up to its 3-dB point, which is known as the natural frequency. This frequency is the same regardless of the order of the filter. However, by increasing the order of the filter, the roll-off in the passband moves closer to its natural frequency and the roll-off in the transition region between the natural frequency and the stopband becomes sharper.

第 3 单元

填空

(1) layers, via, while, discrete components, hundreds of thousands, etched circuit, copper foil, fiberglass, negative image, after, similar process, on

(2) electrical pulses, in, are, to, as well as, communication protocols, graphical interfaces, (engineers, programmers and consultants), implies a structure, (voltage levels, frequencies and duration), hardware, read, transmitted, activate

英汉互译

1) 定制门阵列逻辑、专用集成电路、球栅阵列、多芯片模块和亚纳米数字器件的使用为电磁兼容工程师提供了新的挑战性机遇。

2) 工作在高频的电阻表现为电感和并联的电阻、电容之间的级联。工作在高频的电容表现为电感、电阻和电容的级联。

3) SPICE(侧重集成电路的仿真程序)是一个广泛用于模拟电路和混合电路仿真的程序。SPICE 解决了多组建立在频域、稳态和时域的非线性方程,从而能够对晶体管和门电路的设计行为进行模拟。

4) IBIS(输入/输出缓冲规范)是一种用来对集成电路输入/输出模拟特性进行定义的格式。IBIS 模型是 ASCII 文件。在不泄漏电路设计专利的前提下,IBIS 模型提供了对器件建模所需的行为信息。

5) 在多导线系统中,导线间的过量耦合(即“串扰”)会引发两种有害效应。首先,串扰改变了总线中传输线的有效特性阻抗和传输速度,从而使总线传输线的性能发生改变。此外,串扰将噪声感应到其他导线上。串扰现象的这些特点使得系统性能严重地依赖于数据模式、导线间距和切换速率。

6) Crystal is a transparent quartz material that contains a uniform arrangement of molecules.

7）As high-speed digital systems become more complex，it is becoming increasingly difficult to manage timings and signal integrity.

8）Crosstalk is the coupling of energy from one line to another. It will occur whenever the electromagnetic fields from different structures interact.

9）In digital designs the occurrence of crosstalk is very widespread. Crosstalk will occur on the chip，on the PCB board，on the connectors，on the chip package，and on the connector cables.

10）Many systems operate at high frequencies at which conductors no longer behave as simple wires，but instead exhibit high-frequency effects and behave as transmission lines. If these transmission lines are not handled properly，they can unintentionally ruin system timing.

第 4 单元

填空

（1）share a frequency channel，digitized speech，a number of time slots，data rate，as well as，Apart from，synchronization，from，to，in，in，from，by，there are several frequency channels used in a cell，or，are referred to as，respectively，as that

（2）linear modulation，also known as，over，spectral density of the signal，as，at the time of，and，using CDMA schemes，for，this，in multipath situations，because of，asynchronous，not necessarily，It should be noted that，other，in，signals transmitted by the base station to other mobiles

英汉互译

1）不同的无线射频链路以不同的方式共用频谱带宽。这些共用方式称为“多址模式”。“频分多址”（FDMA）、“时分多址”（TDMA）、“码分多址”（CDMA）和“空分多址”（SDMA）是四种主要的多址模式。

2）移动电话系统的服务区划分为一个个“蜂窝”小区。每个“蜂窝”小区都有一座基站；通过在无线链路上发送、接收信号，基站和“蜂窝”内移动用户通信。基站向移动用户发送信号称作“前向链路”或“下行链路”。移动用户向基站发送信号称作“反向链路”或“上行链路”。每个基站都要和“移动交换中心”（MSC）联络，而 MSC 的作用是把来往于基站的呼叫和其他蜂窝的移动用户、公共交换电话网连接起来。

3）基站使用两类无线信道和移动用户进行通信：传送控制信息的“控制信道”和传送消息的“通信信道”。

4）在 FDMA 方式中，可用频谱划分为若干具有一定带宽的频道；每个呼叫都使用不

同频道。第一代蜂窝系统采用的都是 FDMA 方式。

5）在 TDMA 和 FDMA 系统中，为了防止信道间干扰，某个蜂窝小区使用的频道不会被邻近蜂窝小区使用。在 CDMA 系统中，可以在临近蜂窝小区中使用相同频率，从而增加了系统容量。

6）The generic term channel is normally used to denote a frequency in FDMA system，a time slot in TDMA system，and a code in CDMA system or a combination of these in a mixed system.

7）In FDMA and TDMA systems once a channel（frequency and/or time slot）is allocated to a user，that channel cannot be used during nonactivity periods.

8）The cellular concept was invented by Bell Laboratories and the first commercial analog voice system was introduced in Chicago in October 1983.

9）The third-generation systems aim to provide a seamless network that can provide users voice，data，multimedia，and video services regardless of their location on the network.

10）In third-generation communication systems，satellites are going to play a major role in providing global coverage.

第 5 单元

填空

（1）With the advent of VLSI in the 1980s，standard ICs，of，implement，logic functions，custom ICs，a smaller number of components，into，reduce cost and improve reliability，in，to，If you need a large amount of memory，in，to，to

（2）a variety of chips，in，by，known as，to，depending on，which，each time，to，the most common PLD chips，Unlike gate arrays，in the field，programmable storage chips，rather than

英汉互译

1）有些数字逻辑集成电路和模拟集成电路（如模数转换器）是标准器件或者标准集成电路。可以从产品目录和数据手册当中选择标准集成电路，从经销商处购买。系统制造商和设计人员可以在各种不同的微电子系统（“微电子系统”是指应用微电子技术和集成电路的系统）中使用相同的标准器件。

2）现代亚微米 CMOS 工艺的复杂度和亚微米双极性工艺、BiMOS（即双极性和 CMOS 的组合）工艺是一样的。不过，CMOS 集成电路现已居于主导地位——其产量比采用其他技术制造的集成电路都大。由于规模带来的成本节约，CMOS 集成电路的成本

要低于同等功能的双极性或 BiMOS 集成电路。在某些特殊应用中，仍要使用双极性集成电路和 BiMOS 集成电路。例如，和 CMOS 集成电路相比，双极性集成电路往往能够处理更高的电压。因此，在电力电子学、汽车、电话电路等领域，双极性集成电路和 BiMOS 集成电路是非常有用的。

3）“专用集成电路”不是微处理器这样的通用芯片，而是一种为特殊用途设计的定制芯片。和通用 CPU 相比，使用专用集成电路可以提高性能。因为专用集成电路是使用“硬连线”完成具体工作的，并不需要为取指令和解释指令付出开销。专用集成电路芯片能够以尽可能快的速度完成一项电子操作；当然，其电路设计应该具有高效的结构。

4）“软核”是在可编程逻辑芯片上或在微控器、片上系统的可编程部分实现的。根据生产商的不同，逻辑功能可能以原理图、网表或者硬件描述语言代码的形式提供。软核可以用芯片上的硬核来实现。

5）“硬核”是一种在芯片电路级设计的、具有特定功能的逻辑模块。微处理器就是典型的硬核，微处理器和特定芯片制造商的半导体生产工艺是联系在一起的。微控器或片上系统可能全部都是硬核，也可能是由硬核（硬连线）和软核（可编程）组合而成。

6）EDA（Electronic Design Automation）uses the computer to design and simulate the performance of electronic circuits on a chip.

7）Hardware Description Language is a language used to describe the functions of an electronic circuit for documentation，simulation or logic synthesis.

8）Logic synthesis is the conversion of a high-level electronic circuit description into a list of logic gates and their interconnections，called the“netlis”. Every logic synthesis program understands some subset of Verilog and VHDL.

9）CPLD (Complex PLD) is a programmable logic device that includes a reprogrammable interconnect between the logic blocks. CPLDs are mostly EEPROM and flash based.

10）FPGA is a programmable logic chip with a high density of gates. Containing up to hundreds of thousands of gates，there are a variety of architectures. Some are very sophisticated，including not only programmable logic blocks，but programmable interconnects and switches between the blocks. FPGAs are mostly reprogrammable（EEPROM or flash based）or dynamic（RAM based）.

第 6 单元

填空

（1）analog system，continuous，discrete points，sampling rate，twice，frequency component，is referred to as，a given set of data ，before new data arrives，at a given clock rate ，odd ，because，the amount of data

（2）status bit，makes it difficult to develop real-time systems，efficient，clock cycles，event-driven interrupt programming，real-world signals，respond，meet all of the following demands，multiple，keep up with，a frame of data，serial port，stops what it is doing ，interrupt vector，on ，To transmit data ，transfer a value，a program segment ，or ，to ，precision and flexibility

英汉互译

1）“提取信息”是对现实世界信号进行处理的主要原因。通常，信息的存在形式是信号幅度（绝对幅度或相对幅度）、频率（或频谱成分）、相位（或者和其他信号之间的时序关系）。一旦把有用信息从信号中提取出来，就能以多种方式使用这些信息。

2）有的时候，人们希望对信号中的信息赋以新的格式。语音信号在频分多址电话系统中的传输就是这种情况。模拟技术用于在频谱范围内对语音信道进行“堆叠”，并通过微波中继、同轴电缆或光纤进行传输。在数字传输链路中，首先使用模数转换器将模拟语音信息转化为数字信号。代表单路语音信号的数字信息在时域进行多路复用（时分多址），并通过一条串行数字链路进行传输（如 T 载波系统）。

3）信号处理还可以满足以下需求：在不损失重要信息的前提下，对信号的频率成分进行压缩、赋以新的格式并以低数据率进行传输，从而减小可用带宽。和数字移动射频系统、MPEG 录放设备和高清晰度电视一样，在高速调制解调器和“自适应差分脉冲调制系统”（ADPCM）中也大量使用数据压缩算法。

4）数字信号处理技术源于 20 世纪六、七十年代，当时刚刚出现数字计算机。那时的计算机是相当昂贵的，数字信号处理仅用于某些关键领域，并在雷达和声纳（关乎国家安全）、石油勘探（可以创造巨大利润）、空间探索（数据具有不可替代性）和医学图像（可用于拯救生命）这四个方面获得了开拓性进展。

5）“采样”是模拟信号向数字信号转换的第一步。通过使用“采样保持电路”可以实现“采样”——“采样保持电路”从信号中获取一个样本，并保持其稳定到下一个采样时刻。样本通常是按照固定时间间隔（称作“采样周期”）采集的。假如采样率不够高，就会出现一种称作“混叠”的信号失真。“采样保持电路”的输出传送到“模-数转换器”，转换器会选择一个最接近信号实际幅度的“量化电平”。这就是“模数转换”的第二步。可供选择的电平数取决于转换器的比特数——对于 N-bit 转换器，可用电平数为 2^N。在数字系统中，模拟幅度是无法得到完美表示的，所以就会出现“量化误差”。

6）Digital signal processing is an essential element of countless home and business systems. Its domain can only increase as time proceeds. Thus，DSP is becoming an essential area of expertise for technologists and engineers.

7）Signals，like sound，light，or voltage，are information-bearing variations. Analog signals are real-world signals. They are defined at every point in time and may

take an infinite number of possible amplitudes. Analog signals are not well suited to processing by computer. They may be converted to digital signals through sampling and quantization. Digital signals, on the other hand, are defined only at sampling points, and may take only a finite number of discrete amplitudes. After processing, a digital signal is converted back into an analog signal.

8) Digital Signal Processing can be divided into two categories, fixed point and floating point. These refer to the format used to store and manipulate numbers within the devices. Fixed point DSPs usually represent each number with a minimum of 16 bits; floating point DSPs typically use a minimum of 32 bits to store each value.

9) DSPs are programmed in the same languages as other scientific and engineering applications, usually assembly or C. Programs written in assembly can execute faster, while programs written in C are easier to develop and maintain. In traditional applications, such as programs run on personal computers, C is almost always the first choice. If assembly is used at all, it is restricted to short subroutines that must run with the utmost speed. However, DSP programs are different from traditional software tasks in two important respects. First, the programs are usually much shorter, say, one hundred lines versus ten thousand lines. Second, the execution speed is often a critical part of the application.

第 7 单元

填空

(1) identify the direction of the sound, between, in, frequencies above about 1 kHz, than, on the opposite side of the head, slightly later than, due to its greater distance from the source, Based on, with, coming from the center of the listener's head, While, from, to, This is because, that can provide this information, as, by, from, By measuring the interval between transmission and echo, with, echo localization

(2) to, fidelity, between, To accomplish these two goals, analog wave, a device called an analog-to-digital converter, by, is, as long as, be very similar to, at, that, are stored as bytes, A CD can store up to 74 minutes of music

英汉互译

1) 通常认为,人类听力范围在 20 Hz 和 20 kHz 之间,但人耳对 1 kHz 和 4 kHz 之间的声音要敏感得多。举例来说,收听者能觉察到声压级为 0 dB 的 3 kHz 声音,而 100 Hz 声音却需要具有 40 dB 的声压级(幅度增至 100 倍)。收听者可以辨别 3 kHz 频率处大于

0.3%左右的频差。在 100 Hz 频率处，可分辨频差要达到 3%。可以比较一下，钢琴相邻两键的频差约为 6%。

2）Octave（“八度音阶”或“倍频程”）是指频率上的两倍关系。在钢琴上，一个“八度音阶”包括八个白色琴键，这就是 octave 得名的原因（在拉丁文中，octo 代表“八”）。换句话说，每隔七个白色琴键之后，钢琴的频率就加倍，而整个键盘跨越了七个多一点的“八度音阶”。通常认为人类听力范围在 20 Hz 到 20 kHz 之间，这两个频率分别对应于钢琴键盘左边约 1/2“八度”处的频率和钢琴键盘右边两个“八度”处的频率。

3）“压扩”的实现方法有三种：(1)先用非线性电路对模拟信号进行处理，然后输出到 8 位线性 ADC 。(2)使用内部采用非均匀量化间隔的 8 位 ADC。(3)使用 12 位线性 ADC，之后使用一张“12 位进－8 位出”的数字查找表。这三种方法所需的非线性处理是一样的，只不过这种非线性处理出现在不同的地方：模拟电路中、ADC 中或者数字电路中。

4）人类对连续声音（如某件乐器发出的音符）的感知往往被划分为三个部分：响度、音调和音品。“响度”度量的是声波强度。“音调”是声音中的基波频率。虽然这两个感知参量起到的某些作用不好理解，但是却和易于描述的物理量之间存在着明确的关系。而“音品”就比较复杂，它是由信号中的谐波分量决定的。

5）语音合成和语音识别技术几乎都是基于人类语音产生模型的。大多数人类语声可分为浊音和摩擦音。当空气由肺压出、并经由声带、从口/鼻呼出时，就产生了浊音。声带是沿气流截面拉伸的两片薄片组织，声带正位于喉结之后。声带以 50～1000Hz 频率的振动来响应不断变化的肌肉拉力，这样形成了周期性短促气流注入人的喉咙。元音就属于浊音的一种。

6）An audio signal is a one-dimensional acoustic wave. When an acoustic wave enters the ear, the eardrum vibrates, causing the tiny bones of the inner ear to vibrate along with it, sending nerve pulses to the brain. These pulses are perceived as sound by the listener.

7）Music is a case of audio signal, but a very important one. Another important special case is speech. Human speech tends to be in the 600-Hz to 6,000-Hz range. Speech is made up of vowels and consonants, which have different properties.

8）With 44,100 samples/sec of 16 bits each, an audio CD needs a bandwidth of 705.6 kbps for monaural and 1.411 Mbps for stereo. While this is lower than what video needs, it still takes almost a full T1 channel to transmit uncompressed CD quality stereo sound in real time.

9）Pulse code modulation, as used within the telephone system, uses 8-bit samples made 8,000 times per second. In North America and Japan, 7 bits are for data and 1 is for control; in Europe all 8 bits are for data. This system gives a data rate of 56,000 bps

or 64,000 bps. With only 8,000 samples/sec, frequencies above 4 kHz are lost.

10) Digitized sound can be easily processed by computers in software. Dozens of programs exist for personal computers to allow users to record, display, edit, mix, and store sound waves from multiple sources. Virtually all professional sound recording and editing are digital nowadays.

第 8 单元

填空

(1) To understand video, the two-dimensional image, a function of time, electron beam, the light intensity, called a frame, reconstruct the image, vary, scan lines, In both systems, is turned off, to, has, Rather than, Instead of, in order, that, at

(2) as, that, in, for, the three additive primary colors, a linear superposition of red、green and blue, must be combined into a single composite signal, various methods for displaying color, have, incompatible, existing, RGB is also not the most efficient scheme

英汉互译

1) 通过人眼成像系统,三维世界成像在视网膜上。视网膜上遍布着光接收细胞,这些细胞对 400～700nm 的光有反应。在成像系统当中,我们用镜头和光敏器件组成照相机来模拟人类视觉观察世界的方式。

2) 在计算机领域,显示器用图像宽、高方向上的像素数进行描述。常规电视机表示为 644×483,这意味着 483 条图像线。不过,显示系统都包含一些用于扫描的开销;因此,常规视频扫描的总线数一定大于 483。

3) 传统上,视频扫描系统是由其总线数(包括同步、消隐开销)、帧速率(以 Hz 为单位)和扫描方式(隔行扫描的 2∶1 或逐行扫描的 1∶1)来表征的。在北美和日本,使用 525/59.94/2∶1 扫描系统,其视频模拟带宽约为 5.5 MHz。在欧洲和亚洲,使用 625/50/2∶1 扫描系统,其模拟带宽约为 6.5 MHz。1125/60/2∶1 扫描系统用于高清晰度电视中,其模拟带宽约为 30 MHz。

4) 视频系统采用一个亮度分量和两个色彩分量的形式传送图像数据。亮度分量的传输是很重要的;因为在传输、处理和存储过程中引入的噪声也会引起和亮度相似的感知效应,而且覆盖整个“从黑到白”的范围。

5) 尽管视网膜的形状大致是球体的一部分,但从局部解剖角度讲却是二维的。出于实际原因,我们在照相机中使用了“平面”像平面,而没有使用“球面”像平面。成像系统理论关注的是对像平面上能量连续分布情况的分析。在光学照相机的像平面处有光照下发生化学变化的胶片。光学胶片的有效成分是含在一层由大小和形状仔细控制的微粒组成

的薄层当中。如果微粒的密度足够大，就可以重现出具备很强原始场景感的图像。在胶片介质中的微粒越细小、排列密度越大，胶片记录空间细节的能力就越强。

6) The human eye has the property that when an image appears on the retina, the image is retained for some number of milliseconds before decaying. If a sequence of images is drawn line by line at 50 images/sec, the eye does not notice that it is looking at discrete images. All video systems exploit this principle to produce moving pictures.

7) All compression systems require two algorithms: one for compressing the data at the source, and another for decompressing it at the destination. In the literature, these algorithms are referred to as the encoding and decoding algorithms, respectively.

8) The JPEG (Joint Photographic Experts Group) is used for compressing continuous-tone still pictures (e. g., photographs). JPEG was developed by photographic experts working under the joint auspices of ITU, ISO, and IEC.

9) Aspect ratio is the ratio of image width to height. Conventional television has an aspect ratio of 4∶3. High-definition television uses a wider ratio of 16∶9. Cinema commonly uses 1.85∶1 or 2.35∶1.

10) MPEG stands for Moving Picture Experts Group, is the name of family of standards used for coding audio-visual information (e. g., movies, video, music) in a digital compressed format. The major advantage of MPEG compared to other video and audio coding formats is that MPEG files are much smaller for the same quality. This is because MPEG uses very sophisticated compression techniques.

第 9 单元

填空

(1) general-purpose, or, Such systems generally use microprocessors, custom-designed chips, as well as, operating systems, low-cost consumer products, very limited storage for instructions, In such cases, to, loads its programs into RAM, are called "embedded computers"

(2) a single program, take place concurrently, transaction or message, for a multithreaded program to achieve true performance gains, in, to, calculation-intensive, symmetric, at, audio and video, reentrant code, so that

英汉互译

1) “硬件”就是机器和设备(如处理器、磁盘、磁带、调制解调器和电缆等)。计算机在运行时既需要硬件，也需要软件，没有了哪一个都不行。硬件设计指定了计算机能够遵循

的指令，而指令告诉计算机该做什么。

2）“固件”是无需电源而能保存内容的存储芯片，包括 ROM、PROM、EPROM 和 EEPROM。在存储了程序代码之后，“固件”就成为一种 “硬的软件”。

3）“软件”涉及了“逻辑和语言”。“软件”要对不断变化的事件进行处理，它必须遵循某种逻辑方式。“语言”用来编写软件。分析和编程中的 “逻辑和语言” 往往比明确存储和传输需求更为复杂。

4）“软件包”是一种为向公众出售而开发的应用程序。“软件包”通常是为了满足大量用户而设计的。尽管通过对各种参数进行选择设定，程序也可以做到适合某个用户的喜好，但它不如定制软件更具个性。

5）“进程”是指计算机上某个用户一段特定代码的执行过程。假如爱丽丝在 8:30 运行微软 WORD 程序，而在 8:36 再次运行，那么她产生了两个进程。如果爱丽丝和鲍伯在 8:30 在同一台计算机上运行 WORD 程序，他们也产生了两个进程。

6）A real-time system is a computer system that responds to input signals fast enough to keep an operation moving at its required speed.

7）Real-Time Operating System is an operating system designed for use in a real-time computer system.

8）System software is a program used to control the computer and develop and run application programs. It includes operating systems，TP monitors，network operating systems and database managers.

9）Hardware is“storage and transmission”. The more memory and disk storage a computer has，the more work it can do. The faster the memory and disks transmit data and instructions to the CPU，the faster it gets done. A hardware requirement is based on the size of the databases that will be created and the number of users or applications that will be served at the same time.

10）The software is instructions for the computer. A series of instructions that performs a particular task is called a“program”. The two major categories of software are“system software”and“application software”. System software is made up of control programs such as the operating system and database management system（DBMS）. Application software is any program that processes data for the user（spreadsheet，word processor，etc.）. A common misconception is that software is data. It is not. Software tells the hardware how to process the data. Software is“run”. Data are“processed”.

第 10 单元

填空

（1）when，on，from，reference point，a high voltage，including，rather than，taking accurate measurements，share，integrated circuits，conduction paths，static electricity，leads，charges

（2）design verification，oscilloscope or logic analyzer，to，partner，stimulus，make up，device-under-test，real-world signals，industry-standard tool，can，for，Similarly，their designs' performance，with，from，a complete measurement solution

英汉互译

1）“波”可被定义为每隔一段时间重复出现的、由变化数值构成的曲线。在自然界中，波是很常见的——如声波、脑电波、海浪、光波和电压波形等——这都是周期性出现的现象。信号源通常产生重复出现可控的电气波形。

2）波形具备许多特性，而其关键属性是和幅度、频率、相位相关的。波形的幅度特性、频率特性和相位特性是信号源用来优化波形的。

3）“数字多用表”是一种用来测量电学量（电压、电流、电阻）的仪表。“数字多用表”的用途广泛、形状多样、尺寸各异。“模拟多用表”在印制刻度上移动指针来显示结果。虽然指针移动有利于显示渐进的变化，但精确度却不如数字显示。“数字多用表”在数字液晶读出器或者发光二极管读出器上显示精确的测量结果。

4）“示波器”是一种在屏幕上显示电信号（波形和脉冲）的测量仪器。“示波器”能够产生“时基”用来进行信号测量。

5）在测量应用中，选择合适的探头是获得最佳信号保真度的关键。在高频测量中，采用有源探头可以获得更加真实的信号再现效果和保真度。

6）Not all signals have their amplitude variations centered on a ground（0 V）reference. The“offset”voltage is the voltage between circuit ground and the center of the signal's amplitude.

7）Differential signals are those that use two complementary paths carrying copies of the same signal in equal and opposite polarity（relative to ground）.

8）Rise and fall time are measures of the time it takes the signal edge to make a transition from one state to another. In modern digital circuitry，these values are usually several nanoseconds.

9）Bandwidth means the transmission capacity of an electronic pathway such as a

communications line or computer bus. In a digital line, it is measured in bits per second or bytes per second. In an analog line, it is the difference between the highest and lowest frequencies and is measured in Hertz .

10) The noise is an extraneous signal that invades an electrical transmission. It can come from strong electrical or magnetic signals in nearby lines, from poorly fitting electrical contacts, and from power line spikes.

缩略语表

(Acronyms & Abbreviations)

A

A/D	analog to digital
ac	alternating current
ACF	auto-correlation function
ACK	acknowledge
ADC	analog-to-digital converter
ADPCM	adaptive differential pulse-code modulation
ADSL	asynchronous/asymmetrical digital-subscriber loop line
AES	Advanced Encryption Standard
AF	audio frequency
AFC	automatic frequency control
AFE	analog front end
AGC	automatic gain control
AGP	accelerated graphics port
AHDL	analog hardware-description language
AI	artificial intelligence
ALU	arithmetic-logic unit
AM	amplitude modulation
ANSI	American National Standards Institute
APD	avalanche photodiode
API	application-programming interface
ASCII	American National Standard Code for Information Interchange
ASIC	application-specific integrated circuit
ASR	automatic speech recognition
ASSP	application-specific standard product
ATE	automatic test equipment
ATELP	adaptive-transform-excited-linear prediction

ATM	asynchronous transfer mode (communications)
ATSC	Advanced Television Systems Committee
ATV	advanced television (high-definition television)
avg	average
AWG	American wire gauge (standard)
AWG	arbitrary-waveform generator

B

B-CDMA	broadband code division multiple access (communications)
Basic	beginners all-purpose symbolic instruction code (software)
BBS	bulletin board system
BCD	binary-coded decimal
BER	bit error rate
BFSK	binary frequency-shift keying
BGA	ball-grid array (IC package)
BiCMOS	bipolar complementary metal-oxide semiconductor
BIOS	basic input/output system (computer)
BJT	bipolar junction transistor
BMP	bit-mapped (graphics file format)
BPF	band-pass filter
BPSK	binary phase-shift keying

C

CAD	computer-aided design
CAE	computer-aided engineering
CAM	content-addressable memory/ computer-aided manufacturing
CAN	controller-area network
CAT	computer-aided test
CBGA	ceramic ball-grid array (IC package)
CCD	charge-coupled device
CCF	cross-correlation function
CCS	common channel signalling (telecommunications)
CCTV	closed-circuit TV
CDMA	code division multiple access (digital communications standard)
CD-R	CD-recordable

CELP	code-excited linear prediction (speech coding)
CISC	complex-instruction-set computing
CMOS	complementary metal-oxide semiconductor
CMR	common-mode ratio
CMRR	common-mode rejection ratio
CMVR	common-mode voltage ratio
CMY	cyan, magenta, yellow (printing colors)
codec	coder-decoder
CPLD	complex programmable-logic device
CRC	cyclic redundancy check
CRT	cathode-ray tube
CSMA/CD	carrier-sense multiple access/collision detection (LAN)
CW	continuous wave

D

D/A	digital-to-analog
DAB	digital-audio broadcasting
DAC	digital-to-analog converter
DAS	data-acquisition system
DAT	digital audio tape
dB	decibel
DBPSK	differential binary phase-shift keying (communications)
dc	direct current
dc/dc	dc to dc
DCT	discrete cosine transform
DDR	double data rate
DDS	direct digital synthesis
DECT	Digital European Cordless Telephone (telecommunication standard)
DES	Data Encryption Standard
DFT	discrete Fourier transform
DIP	dual in-line package(IC package)
DMA	direct-memory access
DMT	discrete multitone (modulation)
DNL	differential nonlinearity (ADC performance)
DNS	domain-name system (Internet)

dpdt	double-pole，double-throw
dpi	dots per inch
dpst	double-pole，single-throw
DQPSK	differential quadrature phase-shift keying
DRAM	dynamic random-access memory
DSL	digital subscriber loop
DSO	digital-storage oscilloscope /digital-sampling oscilloscope
DSP	digital signal processing/processor
DSSS	direct-sequence spread spectrum (communication)
DTE	data-termination equipment
DTV	digital television
DUT	device under test
DVD	digital versatile disk
DVI	digital video interactive
DVM	digital voltmeter
DWDM	dense-wavelength-division multiplexing
DWT	discrete wavelet transform

E

ECC	error-correction code
ECL	emitter-coupled logic
EDA	electronic-design automation
EEPROM	electrically erasable PROM
EIA	Electronic Industries Association (Washington)
EMC	electromagnetic compatibility
EMF	electromagnetic field
EMI	electromagnetic interference
ENOB	effective number of bits (ADC performance)
EPLD	erasable programmable-logic device
ESD	electrostatic discharge
EUT	equipment under test

F

FAT	file-allocation table (digital disk format)
FC	fibre channel (communication)

FDDI	fibre distributed data interface (communications)
FDMA	frequency-division multiple access (communications)
FDX	full-duplex (communications)
FEC	forward error correction
FET	field-effect transistor
FFT	fast Fourier transform
FHSS	frequency hopping spread spectrum (communications)
FIFO	first in, first out (memory)
FIR	finite-impulse response (filter)
FLOPS	floating-point operations per second
FM	frequency modulation
FMCW	frequency-modulated continuous wave (radar)
FPGA	field-programmable gate array
FQFP	fine-pitch quad flat pack(IC package)
FR	full rate (communications)
FRAM	ferroelectric random-access memory
FRED	fast-recovery epitaxial diode (rectifier)
FS	full scale
FSK	frequency-shift keying (communications)
FSM	finite state machine
FTP	File Transfer Protocol (Internet)

G

GCA	gain-controlled amplifier
GFLOPS	giga floating-point operations per second
GIF	graphic image file
GMSK	Gaussian-filtered minimum-shift keying
GOPS	giga-operations per second
GPIB	general-purpose interface bus
GPRS	General Packet Radio Service
GPS	Global Positioning System/Satellite
GSM	Global System for Mobile communications
GUI	graphical user interface

H

HDD	hard-disk drive
HDL	hardware-description language
HDTV	high-definition digital television
HDX	half duplex (communications)
HLL	high-level language (software)
HR	half-rate (communications)
HSL	hue, saturation, luminescence (color graphics)
HSTP	High Speed Transport Protocol
HTML	Hypertext Markup Language (WWW-Internet)
HTTL	high-power transistor-to-transistor logic
HTTP	Hypertext Transfer Protocol

I

I/O	input/output
IC	integrated circuit
ICE	in-circuit emulator
IDE	intelligent drive electronics (disk)
IDFT	inverse discrete Fourier transform
IEE	Institution of Electric Engineers (UK)
IEEE	Institute of Electrical and Electronic Engineers
IF	intermediate frequency
IFFT	inverse fast Fourier transform
INL	integral non-linearity (ADC performance)
IP	intellectual property/ Internet Protocol
IR	infrared
IrDA	Infrared Data Association
IRQ	interrupt request (software)
ISA	industry-standard architecture
ISDN	integrated services digital network (communications)
ISO	International Standards Organization
ISP	in-system programmable/ Internet-service provider
ITU	International Telecommunications Union

J

JEDEC	Joint Electronic Device Engineering Council
JFET	junction field-effect transistor
JPEG	Joint Photographic Experts Group (standard)
JTAG	Joint Test Action Group

K

kHz	kilohertz

L

LAB	logic array block
LAN	local-area network
laser	light amplification by stimulated emission of radiation
LCD	liquid-crystal display
LCR	inductance-capacitance-resistance (meter)
LED	light-emitting diode
LF	low frequency (30 to 300 kHz)
LIFO	last-in, first-out
lm	lumen
LNA	low noise amplifier
LO	local oscillator
LPF	low-pass filter
LSB	least significant bit
LSI	large scale integration
LUT	look-up table
LVDS	low-voltage differential signalling

M

MAC	multiply-accumulate/media-access control
MCU	microcontroller unit
MD	minidisk
MDCT	modified discrete cosine transform
MEMS	microelectromechanical systems
MHz	megahertz

MIDI	musical-instrument digital interface (multimedia)
MIMD	multiple-instruction, multiple-data
MIPS	million instructions per second
MIS	management information system
MMDS	multichannel multipoint distribution service
MMIC	monolithic microwave integrated circuit
MMX	multimedia extension
modem	modulator-demodulator (telecommunications)
MOPS	million operations per second
MOS	metal-oxide semiconductor
MOSFET	metal-oxide semiconductor field-effect transistor
MPEG	Moving Picture Experts Group (standard)
MRI	magnetic resonance imaging
MSB	most significant bit
MSK	minimum shift keying (communications)
MTD	moving-target detection (radar)
mux	multiplexer

N

NCO	numerically controlled oscillator
NMOS	N-channel metal-oxide semiconductor
NOP	no-operation instruction
NOS	network operating system
NRE	nonrecurring engineering
NRZ	nonreturn to zero
NTSC	National Television Systems Committee (standard)
NVRAM	non-volatile random-access memory
NWDM	narrow-wavelength-division multiplexing

O

OC	optical carrier
OFDM	orthogonal frequency division multiplexing (communications)
OOP	object-oriented programming
OQPSK	offset-quadrature phase-shift keying (communications)
OS	operating system (computer)

OSR	oversampling ratio (ADC)
OTDR	optical time-domain reflectometer

P

p-p	peak-to-peak
PAL	phase-alternation line/ programmable-array logic
PAM	pulse-amplitude modulation
PBX	private branch exchange
PC	personal computer
PCB	printed-circuit board
PCI	Peripheral Component Interconnect
PCM	pulse-code modulation
PCMCIA	Personal Computer Memory Card International Association
PD	pulsed Doppler (radar)
PDA	personal digital assistant (portable computing)
PDP	plasma display panel
PFC	power-factor correction (power supply)
PGA	pin-grid array(IC package)
PHY	physical layer
pixel	picture element
PLD	programmable-logic device
PLL	phase-locked loop
PM	phase modulation
PMOS	p-type metal-oxide semiconductor
pnp	p-type n-type p-type
POST	power-on self tests
PRN	pseudo-random noise
PROM	programmable ROM
PRS	pattern recognition system
PSD	power spectral density
PSK	phase-shift keying
PSPDN	packet switched public data network
PSTN	Public Switched Telephone Network
PWM	pulse-width modulation
PXI	PCI eXtensions for Instrumentation

Q

Q	quality factor (of a resonant circuit)
QAM	quadrature amplitude modulation
QoS	quality of service (communications)
QPSK	quadrature phase-shift keying (communications)

R

RAM	random access memory
RAMDAC	random access memory, digital-to-analog converter (combination)
RF	radio frequency
RGB	red, green, blue
RISC	reduced-instruction-set computer
RLC	resistance-inductance-capacitance
RLE	run-length encoding (data compression)
rms	root-mean-square
ROM	read-only memory
RTOS	real-time operating system
RW	read/write

S

S/H	sample and hold
SAR	synthetic aperture radar
SAW	surface acoustic wave
SCSI	Small Computer System Interface
SDRAM	synchronous DRAM
SDTV	standard-definition digital television
SECAM	Systeme Electronique Couleur Avec Memoire
SFC	Shannon-Farro Coding (data compression)
SFDR	spurious free dynamic range
SHA	sample-and-hold amplifier (ADC)
SHF	super high frequency
SIMD	single-instruction, multiple-data
SINAD	signal-to-noise-and-distortion (ratio)
SMC	surface-mount components

SMD	surface-mount device
SNR	signal-to-noise ratio
SOC	system on chip
SONET	synchronous optical network
SOPC	system on a programmable chip
spdt	single-pole，double-throw
SPI	serial peripheral interface
SPICE	Simulation Program with Integrated Circuit
SPS	samples per second
spst	single-pole，single-throw
SR	set/reset (flip-flop)
SRAM	static random-access memory
SS	spread spectrum (communications)
SSB	single sideband
SSI	small-scale integration
SST	solid state technology
STB	set-top box

T

T/H	track and hold
T/R	transmit and receive
TCM	trellis-coded modulation
TCP/IP	Transmission Control Protocol/Internet Protocol
TDM	time-division multiplexing (communications)
TDMA	time-division multiple access (communications)
TDR	time domain reflectometer
TEM	transverse electromagnetic
TF	transfer function
THD	total harmonic distortion
TIFF	tagged image format file (computer graphics)
TTL	transistor-transistor logic

U

UART	universal asynchronous receiver-transmitter (communications)
uC	microcontroller

UHF	ultrahigh frequency
ULSI	ultra large-scale integration
uP	microprocessor
UPS	uninterruptible power source
USB	Universal Serial Bus
UTP	unshielded twisted pair
UV	ultraviolet

V

V/F	voltage to frequency
VC	virtual channel
VCO	voltage-controlled oscillator
VF	voice frequency (300～3,400 Hz)
VFC	voltage-to-frequency converter
VGA	video-graphics adapter
VHDL	very-high-speed integrated-circuit hardware-description language
VHF	very high-frequency (30 kHz～300 MHz)
VLF	very low frequency (below 30 kHz)
VNA	vector network analyzer
VOD	video-on-demand
VoIP	voice over Internet Protocol
VPN	virtual private network
VQ	vector quantization
VR	virtual reality
VSB	vestigial sideband

W

WAN	wide-area network
WAP	Wireless Application Protocol
WDM	wavelength-division multiplexing
WLAN	wireless local area network
WWW	World Wide Web of the Internet

X

xDSL	Digital Subscriber Line

XML	eXtensible Markup Language

Y

YUV	luminance/chrominance

数学表达式的英文读法

Pronunciation of Mathematical Expressions

逻辑

$\exists$	there exists
$\forall$	for all
$p\Rightarrow q$	p implies q / if p, then q
$p\Leftrightarrow q$	p if and only if q / p is equivalent to q / p and q are equivalent

集合

$x \in A$	x belongs to A / x is an element (or a member) of set A
$x \notin A$	x does not belong to A / x is not an element (or a member)of A
$A\subset B$	A is contained in B / A is a subset of B
$A\supset B$	A contained B / B is a subset of A
$A\cap B$	A meet B / A intersection B / intersection of A and B
$A\cup B$	A join B / A union B
$A\backslash B$	A minus B / the difference between A and B
$A\times B$	A cross B / the Cartesian product of A and B

实数

$x+1$	x plus one
$x-1$	x minus one
$x\pm 1$	x plus or minus one
xy	xy / x multiplied by y
$(x-y)(x+y)$	x minus y, x plus y
$\frac{x}{y}$	x over y
$x=5$	x equals 5 / x is equal to 5
$x\neq 5$	x (is) not equal to 5
$x\equiv y$	x is equivalent to (or identical with) y

$x>y$	x is greater than y
$x\geqslant y$	x is greater than or equal to y
$x<y$	x is less than y
$x\leqslant y$	x is less than or equal to y
$0<x<1$	zero is less than x is less than 1
$0\leqslant x\leqslant 1$	zero is less than or equal to x is less than or equal to 1
$\|x\|$	mod x / modulus x
x^2	x squared / x (raised) to the power 2
x^3	x cubed
x^4	x to the fourth / x to the power four
x^n	x to the n th / x to the power n
x^{-n}	x to the (power) minus n
$\sqrt{x}$	(square) root x / the square root of x
$\sqrt[3]{x}$	cube root (of) x
$\sqrt[4]{x}$	fourth root (of) x
$\sqrt[n]{x}$	n th root (of) x
$(x+y)^2$	x plus y all squared
$\left(\frac{x}{y}\right)^2$	x over y all squared
$n!$	n factorial
$\hat{x}$	x hat
$\bar{x}$	x bar
$\tilde{x}$	x tilde
x_i	xi / x subscript i / x suffix i / x sub i
$\sum_{i=1}^{n} a_i$	the sum from i equals one to n a_i/
	the sum as i runs from one to n of the a_i

线性代数

$\\|x\\|$	the norm (or modulus) of x
$\overrightarrow{OA}$	OA / vector OA
$\overline{OA}$	OA / the length of the segment OA
A^T	A transpose / the transpose of A
A^{-1}	A inverse / the inverse of A

函数

$f(x)$	f x / f of x / the function f of x
$f: S \rightarrow T$	a function f from S to T
$x \mapsto y$	x maps to y / x is sent (or mapped) to y
$f'(x)$	f prime x / f dash x /the (first) derivative of f with respect to x
$f''(x)$	f double-prime x / f double-dash x / the second derivative of f with respect to x
$f'''(x)$	f triple-prime x / f triple-dash x / the third derivative of f with respect to x
$f^{(4)}(x)$	f four x / the fourth derivative of f with respect to x
$\frac{\partial f}{\partial x_1}$	the partial (derivative) of f with respect to x_1
$\frac{\partial^2 x}{\partial x_1^2}$	the second partial (derivative) of f with respect to x_1
$\int_0^\infty$	the integral from zero to infinity
$\lim\limits_{x \rightarrow 0}$	the limit as x approaches zero
$\lim\limits_{x \rightarrow +0}$	the limit as x approaches zero from above
$\lim\limits_{x \rightarrow -0}$	the limit as x approaches zero from below
$\log_e y$	log y to the base e / log to the base e of y / natural log (of) y
$\ln y$	log y to the base e / log to the base e of y / natural log (of) y

参 考 文 献

[1] Steven W. Smith. The Scientist and Engineer's Guide to Digital Signal Processing (Second Edition), California Technical Publishing, San Diego, California, 1999.

[2] Andrew Bateman, Lain Paterso Stephens. The DSP handbook: Algorithms, Applications and Design Techniques, Pearson Education limited, 2002.

[3] Andrew S. Tanenbaum. Computer Networks (4th ed.), Prentice Hall, 2002.

[4] Joyce Vande Vegte. Fundamentals of Digital Signal Processing, Prentice Hall, 2002.

[5] Walt Kester. Mixed signal processing and DSP design, Newns, 2003.

[6] Carl R. Nassar. Telecommunication Demystified: A Streamlined Course in Digital Communications (and some Analog) for EE Students and Practicing Engineers. LLH Technology Publishing, 2001.

[7] Stephen A. Edwards. Languages for Digital Embedded Systems, Kluwer Academic Publishers, 2000.

[8] Mark Zwo linski. Digital System Design with VHDL, Prentice Hall, 2000.

[9] Clive Maxfield. The Design Warrior's Guide to FPGAs: Devices, Tools and Flows, Newnes, 2004.

[10] CRC Press LLC. Handbook of Antennas in Wireless Communications, 2002.

[11] Stephen H. Hall, Garrett W. Hall, James A. McCall. High-Speed Digital System Design—A Handbook of Interconnect Theory and Design Practices, John Wiley & Sons, Inc., 2000.

[12] Charles Poynton. A Technical Introduction to Digital Video,John Wiley & Sons, 1996.

[13] Lewin Edwards. Embedded System Design on a Shoestring: Achieving High Performance with a Limited Budget, Newns, 2003.

[14] Herman Bruyninckx. Real-Time and Embedded Guide, K. U. Leuven Mechanical Engineering, 2001.

[15] Bonnie Baker. Filtering? Before or after?, www.edn.com, 2003.

[16] Stephen Brown, Jonathan Rose. FPGA and CPLD Architecture: A Tutorial. IEEE Design & Test of Computers, 1996.

[17] Michael Barr. "Choosing an RTOS", Embedded Systems Programming, January 2003.

［18］ Maxim Integrated Products，Inc. A Filter Primer，2001.
［19］ Maxim Integrated Products，Inc. Choosing The Right Power Supply IC for Your Application，2001.
［20］ Maxim Integrated Products，Inc. Specifying Quartz Crystal，2000.
［21］ Semiconductor Components Industries，LLC（SCILLC）. SWITCHMODE™ Power Supplies Reference Manual and Design Guide（Rev. 1），Sept. 1999.
［22］ Cypress Semiconductor Corporation. Clock Terminology，1997.
［23］ The Aerospace Corporation. The Global Positioning System，1997.
［24］ Texas Instruments，Inc. Op Amps for Everyone，2002.
［25］ Texas Instruments，Inc. Programmable DSP Platform for Digital Still Cameras. Technical Journal，January-March，2000.
［26］ Tektronix，Inc. XYZs of Signal Sources，2003.
［27］ Tektronix，Inc. An Overview of Signal Source Technology and Applications，2000.
［28］ Tektronix，Inc. XYZs of Oscilloscopes，2001.
［29］ Tektronix，Inc. The XYZs of Logic Analyzers，2001.
［30］ Tektronix，Inc. A Digital Designer's Guide to Verifying Signal Integrity，2002.
［31］ National Instruments Corporation. Virtual Instrumentation，2002.
［32］ http：//www. howstuffworks. com.
［33］ http：//www. wikipedia. org.
［34］ http：//www. arm. com.
［35］ http：//www. reed-electronics. com/ednmag.
［36］ http：//www. ganssle. com/articles. htm.
［37］ http：//www. dolby. com.
［38］ http：//www. webopedia. com.
［39］ http：//www. britannica. com.
［40］ http：//www. m-w. com/cgi-bin/dictionary.

电子信息工程专业英语教程

（第3版）

本书编者在前一版的基础上，吸取多所大学在使用本书过程中提出的诸多宝贵意见，对全书进行修订和补充。本书的主要目的是使读者掌握电子信息工程专业英语术语及用法，培养和提高读者阅读和翻译专业英语文献资料的能力。本书由10个主题单元组成，涵盖电子信息领域的主要技术分支，主要内容包括电子器件、电子电路、电子系统组件、电子系统、现代数字设计、数字信号处理、语音和音频、图像和视频、嵌入式应用、电子仪器与测量等内容。每个主题单元由3篇课文、3篇阅读材料、课文词汇、课文注释和练习组成，在书后还附有课文参考译文、练习参考答案及缩略语表。为了方便教学，本书另配有电子教案、阅读材料参考译文和授课建议，向采纳本书作为教材的教师免费提供。

本书可作为电子信息工程专业的专业英语教材，也可供从事相关专业的工程技术人员参考使用。

策划编辑：杨丽娟
责任编辑：杨丽娟
责任美编：孙焱津

ISBN 978-7-121-14018-1

定价：45.00 元